```
Welcome to STN International!  Enter x:x

LOGINID:SSSPTA1805JXB

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2


* * * * * * * * * *   Welcome to STN International   * * * * * * * * * *

 NEWS  1           Web Page URLs for STN Seminar Schedule - N. America
 NEWS  2  Apr 08   "Ask CAS" for self-help around the clock
 NEWS  3  Apr 09   BEILSTEIN: Reload and Implementation of a New Subject Area
 NEWS  4  Apr 09   ZDB will be removed from STN
 NEWS  5  Apr 19   US Patent Applications available in IFICDB, IFIPAT, and IFIUDB
 NEWS  6  Apr 22   Records from IP.com available in CAPLUS, HCAPLUS, and ZCAPLUS
 NEWS  7  Apr 22   BIOSIS Gene Names now available in TOXCENTER
 NEWS  8  Apr 22   Federal Research in Progress (FEDRIP) now available
 NEWS  9  Jun 03   New e-mail delivery for search results now available
 NEWS 10  Jun 10   MEDLINE Reload
 NEWS 11  Jun 10   PCTFULL has been reloaded
 NEWS 12  Jul 02   FOREGE no longer contains STANDARDS file segment
 NEWS 13  Jul 22   USAN to be reloaded July 28, 2002;
                   saved answer sets no longer valid
 NEWS 14  Jul 29   Enhanced polymer searching in REGISTRY
 NEWS 15  Jul 30   NETFIRST to be removed from STN
 NEWS 16  Aug 08   CANCERLIT reload
 NEWS 17  Aug 08   PHARMAMarketLetter(PHARMAML) - new on STN
 NEWS 18  Aug 08   NTIS has been reloaded and enhanced
 NEWS 19  Aug 19   Aquatic Toxicity Information Retrieval (AQUIRE)
                   now available on STN
 NEWS 20  Aug 19   IFIPAT, IFICDB, and IFIUDB have been reloaded
 NEWS 21  Aug 19   The MEDLINE file segment of TOXCENTER has been reloaded
 NEWS 22  Aug 26   Sequence searching in REGISTRY enhanced
 NEWS 23  Sep 03   JAPIO has been reloaded and enhanced

 NEWS EXPRESS  February 1 CURRENT WINDOWS VERSION IS V6.0d,
               CURRENT MACINTOSH VERSION IS V6.0a(ENG) AND V6.0Ja(JP),
               AND CURRENT DISCOVER FILE IS DATED 05 FEBRUARY 2002
 NEWS HOURS    STN Operating Hours Plus Help Desk Availability
 NEWS INTER    General Internet Information
 NEWS LOGIN    Welcome Banner and News Items
 NEWS PHONE    Direct Dial and Telecommunication Network Access to STN
 NEWS WWW      CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
specific topic.

  All use of STN is subject to the provisions of the STN Customer
  agreement.  Please note that this agreement limits use to scientific
  research.  Use for software development or design or implementation
  of commercial gateways or other similar uses is prohibited and may
  result in loss of user privileges and other penalties.

  * * * * * * * * * * * * * * STN  Columbus  * * * * * * * * * * * * * * *



FILE 'HOME' ENTERED AT 17:48:35 ON 04 SEP 2002

=> file .pub
COST IN U.S. DOLLARS                          SINCE FILE      TOTAL
```

```
                                                      ENTRY     SESSION
FULL ESTIMATED COST                                    0.21        0.21

FILE 'MEDLINE' ENTERED AT 17:48:50 ON 04 SEP 2002

FILE 'BIOSIS' ENTERED AT 17:48:50 ON 04 SEP 2002
COPYRIGHT (C) 2002 BIOLOGICAL ABSTRACTS INC.(R)

=> s sequence and (positive or negative)
L1      107059 SEQUENCE AND (POSITIVE OR NEGATIVE)

=> s l1 and (polynucleotide or polypeptide or DNA or amino acid or nucleotide)
L2       80894 L1 AND (POLYNUCLEOTIDE OR POLYPEPTIDE OR DNA OR AMINO ACID OR
               NUCLEOTIDE)

=> s l2 and (data or compression)
L3       46668 L2 AND (DATA OR COMPRESSION)

=> s l2 and compression
L4          18 L2 AND COMPRESSION

=> duplicate remove l4
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE?  Y/(N):n
PROCESSING COMPLETED FOR L4
L5           15 DUPLICATE REMOVE L4 (3 DUPLICATES REMOVED)

=> d 1-15 bib ab

L5   ANSWER 1 OF 15     MEDLINE                              DUPLICATE 1
AN   2001148483     MEDLINE
DN   21109790   PubMed ID: 11175899
TI   Structure of the N6-adenine DNA methyltransferase M.TaqI in
     complex with DNA and a cofactor analog.
CM   Comment in: Nat Struct Biol. 2001 Feb;8(2):101-3
AU   Goedecke K; Pignot M; Goody R S; Scheidig A J; Weinhold E
CS   Max-Planck-Institut fur molekulare Physiologie, Abteilung Physikalische
     Biochemie, Otto-Hahn-Str. 11, D-44227 Dortmund, Germany.
SO   NATURE STRUCTURAL BIOLOGY, (2001 Feb) 8 (2) 121-5.
     Journal code: 9421566. ISSN: 1072-8368.
CY   United States
DT   Journal; Article; (JOURNAL ARTICLE)
LA   English
FS   Priority Journals
OS   PDB-1G38
EM   200103
ED   Entered STN: 20010404
     Last Updated on STN: 20010404
     Entered Medline: 20010315
AB   The 2.0 A crystal structure of the N6-adenine DNA
     methyltransferase M.TaqI in complex with specific DNA and a
     nonreactive cofactor analog reveals a previously unrecognized
     stabilization of the extrahelical target base. To catalyze the transfer of
     the methyl group from the cofactor S-adenosyl-l-methionine to the 6-amino
     group of adenine within the double-stranded DNA sequence
     5'-TCGA-3', the target nucleoside is rotated out of the DNA
     helix. Stabilization of the extrahelical conformation is achieved by
     DNA compression perpendicular to the DNA helix
     axis at the target base pair position and relocation of the partner base
     thymine in an interstrand pi-stacked position, where it would sterically
     overlap with an innerhelical target adenine. The extrahelical target
     adenine is specifically recognized in the active site, and the 6-amino
     group of adenine donates two hydrogen bonds to Asn 105 and Pro 106, which
     both belong to the conserved catalytic motif IV of N6-adenine DNA
     methyltransferases. These hydrogen bonds appear to increase the partial
```

negative charge of the N6 atom of adenine and activate it for direct nucleophilic attack on the methyl group of the cofactor.

L5 ANSWER 2 OF 15 MEDLINE
AN 2001393986 MEDLINE
DN 21151315 PubMed ID: 11256808
TI Pressure-dependent changes in the structure of the melittin alpha-helix determined by NMR.
AU Iwadate M; Asakura T; Dubovskii P V; Yamada H; Akasaka K; Williamson M P
CS Department of Biotechnology, Tokyo University of Agriculture and Technology, Japan.
SO JOURNAL OF BIOMOLECULAR NMR, (2001 Feb) 19 (2) 115-24.
Journal code: 9110829. ISSN: 0925-2738.
CY Netherlands
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200107
ED Entered STN: 20010716
Last Updated on STN: 20010716
Entered Medline: 20010712
AB A novel method is described, which uses changes in NMR chemical shifts to characterise the structural change in a protein with pressure. Melittin in methanol is a small alpha-helical protein, and its chemical shifts change linearly and reversibly with pressure between 1 and 2000 bar. An improved relationship between structure and HN shift has been calculated, and used to drive a molecular dynamics-based calculation of the change in structure. With pressure, the helix is compressed, with the H-O distance of the NH-O=C hydrogen bonds decreased by 0.021 +/- 0.039 A, leading to an overall **compression** along the entire helix of about 0.4 A, corresponding to a static compressibility of 6 x 10(-6) bar(-1). The backbone dihedral angles phi and psi are altered by no more than +/- 3 degrees for most residues with a **negative** correlation coefficient of -0.85 between phi(i) and psi(i - 1), indicating that the local conformation alters to maintain hydrogen bonds in good geometries. The method is shown to be capable of calculating structural change with high precision, and the results agree with structural changes determined using other methodologies.

L5 ANSWER 3 OF 15 MEDLINE DUPLICATE 2
AN 1999264027 MEDLINE
DN 99264027 PubMed ID: 10333232
TI HPV in situ hybridization with catalyzed signal amplification and polymerase chain reaction in establishing cerebellar metastasis of a cervical carcinoma.
AU Huang C C; Kashima M L; Chen H; Shih I M; Kurman R J; Wu T C
CS Department of Pathology, The Johns Hopkins Medical Institutions, Baltimore, MD, USA.
NC 5 POL 34582-01
SO HUMAN PATHOLOGY, (1999 May) 30 (5) 587-91.
Journal code: 9421547. ISSN: 0046-8177.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199905
ED Entered STN: 19990607
Last Updated on STN: 19990607
Entered Medline: 19990526
AB We report an unusual case of cerebellar metastasis from a cervical adenosquamous carcinoma in which molecular techniques assisted in establishing the correct diagnosis. The patient was a 43-year-old woman with surgically unresectable cervical carcinoma diagnosed 2 years before presenting with neurological symptoms. A magnetic resonance imaging scan showed a large, enhancing cerebellar lesion with significant brain stem

compression. The excised cerebellar tumor resembled a small cell carcinoma and was initially not thought to be a metastasis from the cervical adenosquamous carcinoma. In situ hybridization with catalyzed signal amplification and polymerase chain reactions with primers specific for human papilloma virus (HPV) types 16 and 18 were used to determine the relationship between the cervical and the cerebellar neoplasms. A **positive** signal was present in the nuclei of both neoplasms by in situ hybridization using HPV16/18 **DNA** probes. Polymerase chain reaction revealed the presence of HPV-18 **DNA sequences** in the cervical and cerebellar neoplasms confirming that the cerebellar neoplasm was a metastasis from the cervical primary.

L5 ANSWER 4 OF 15 MEDLINE
AN 1999347891 MEDLINE
DN 99347891 PubMed ID: 10421523
TI New techniques for **DNA sequence** classification.
AU Wang J T; Rozen S; Shapiro B A; Shasha D; Wang Z; Yin M
CS Department of Computer and Information Science, New Jersey Institute of Technology, University Heights, Newark 07102, USA.. jason@cis.njit.edu
SO JOURNAL OF COMPUTATIONAL BIOLOGY, (1999 Summer) 6 (2) 209-18.
Journal code: 9433358. ISSN: 1066-5277.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199908
ED Entered STN: 19990913
Last Updated on STN: 19990913
Entered Medline: 19990831
AB **DNA sequence** classification is the activity of determining whether or not an unlabeled **sequence** S belongs to an existing class C. This paper proposes two new techniques for **DNA sequence** classification. The first technique works by comparing the unlabeled **sequence** S with a group of active motifs discovered from the elements of C and by distinction with elements outside of C. The second technique generates and matches gapped fingerprints of S with elements of C. Experimental results obtained by running these algorithms on long and well conserved Alu **sequences** demonstrate the good performance of the presented methods compared with FASTA. When applied to less conserved and relatively short functional sites such as splice-junctions, a variation of the second technique combining fingerprinting with consensus **sequence** analysis gives better results than the current classifiers employing text **compression** and machine learning algorithms.

L5 ANSWER 5 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1999:104663 BIOSIS
DN PREV199900104663
TI Analyses of secondary structures in **DNA** by pyrosequencing.
AU Ronaghi, Mostafa (1); Nygren, Malin; Lundeberg, Joakim; Nyren, Pal
CS (1) Dep. Biotechnology, Royal Inst. Technol., SE-100 44 Stockholm Sweden
SO Analytical Biochemistry, (Feb. 1, 1999) Vol. 267, No. 1, pp. 65-71.
ISSN: 0003-2697.
DT Article
LA English
AB A common problem in conventional **DNA** sequencing is the occurrence of **DNA sequence compressions** during gel electrophoresis, leading to misreading of the **sequence**. These **compressions** are usually due to secondary structures in the **DNA** fragment. In this study, we present a non-gel-based **DNA** sequencing technique that facilitates analysis of such **DNA** regions. A part of the polymorphic pertussis toxin promoter region in five different Bordetella species was successfully resolved by the new technique. The obtained **sequence** data revealed four related palindromic **sequences**. The ability of different

DNA polymerases to read through such secondary structures is also described.

L5 ANSWER 6 OF 15 MEDLINE DUPLICATE 3
AN 1999071263 MEDLINE
DN 99071263 PubMed ID: 9824357
TI Management of fibrosing pancreatitis in children presenting with obstructive jaundice.
AU Sylvester F A; Shuckett B; Cutz E; Durie P R; Marcon M A
CS Division of Gastroenterology and Nutrition, The Hospital for Sick Children, Toronto, Ontario, Canada.
SO GUT, (1998 Nov) 43 (5) 715-20.
Journal code: 2985108R. ISSN: 0017-5749.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Abridged Index Medicus Journals; Priority Journals
EM 199901
ED Entered STN: 19990202
Last Updated on STN: 19990202
Entered Medline: 19990120
AB BACKGROUND: Children with fibrosing pancreatitis are conventionally treated surgically to relieve common bile duct (CBD) obstruction caused by pancreatic **compression**. Residual pancreatic function has not been formally tested in these patients. AIMS: To evaluate the usefulness of non-surgical temporary drainage in children with fibrosing pancreatitis and to assess pancreatic function after resolution of their CBD obstruction. PATIENTS: Four children (1.5-13 years; three girls). METHODS AND RESULTS: Abdominal sonography and computed tomography revealed diffuse enlargement of the pancreas, predominantly the head. The CBD was dilated due to **compression** by the head of the pancreas. Pancreatic biopsy specimens obtained in three patients showed notable acinar cell atrophy and extensive fibrosis. Cystic fibrosis was excluded. No other cause of pancreatitis was identified. Pancreatic tissue from one patient contained viral **DNA sequences** for parvovirus B19 detected by polymerase chain reaction; serum IgM to parvovirus was **positive**. Three patients had temporary drainage of the CBD and one patient underwent a choledochojejunostomy. Serial imaging studies revealed resolution of the CBD obstruction with reduction in pancreatic size. Exocrine pancreatic function deteriorated. Three patients developed pancreatic insufficiency within two to four months of presentation. The fourth patient has notably diminished pancreatic function, but remains pancreatic sufficient. None has diabetes mellitus. CONCLUSIONS: Temporary drainage of the CBD obstruction is recommended in fibrosing pancreatitis in children along with close monitoring of the clinical course, before considering surgery.

L5 ANSWER 7 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1998:206408 BIOSIS
DN PREV199800206408
TI Electrophoresis of **DNA** sequencing fragments at elevated temperature in capillaries filled with poly(N-acryloylaminopropanol) gels.
AU Lindberg, Peter; Righetti, Pier Giorgio; Gelfi, Cecilia; Roeraade, Johan (1)
CS (1) Royal Inst. Technol., Dep. Anal. Chem., S-100 44 Stockholm Sweden
SO Electrophoresis, (Dec., 1997) Vol. 18, No. 15, pp. 2909-2914.
ISSN: 0173-0835.
DT Article
LA English
AB The performance of poly(N-acryloylaminopropanol) (poly AAP) get columns, proved to be stable during electrophoresis at elevated temperature, was investigated. The column manufacturing procedure included the preparation of a coating of the inner wall of the fused silica capillary column with linear poly(AAP). Then, a mixture of the AAP monomer, the cross-tinker dihydroxyethylenebisacrylamide (DHEBA) and linear poly(AAP) was introduced

into the column and in situ polymerized (for preparation of linear gel columns, the addition of DHEBA was omitted). The poly(AAP) columns were first evaluated by electrophoresis of oligonucleotides at room temperature and at 50degreeC, utilizing 260 nm UV-absorbance detection. In a further evaluation of column performance, samples of T-terminated **DNA** Sanger fragments from the bacteria Moracella were separated at 200 V/cm electrical field strength, utilizing a 488 nm argon ion laser and a confocal optical setup for laser-induced fluorescence (LIF) detection. A temperature increase from 25degreeC to 50degreeC effectively released a **compression** of **DNA** bands. However, for cross-linked poly(AAP) gel columns, the elevated temperature resulted in a considerable reduction of the **DNA sequence** reading length. When a linear poly(AAP) column was utilized, no detrimental effect of elevated temperature on the separation could be observed.

L5 ANSWER 8 OF 15 MEDLINE
AN 95337322 MEDLINE
DN 95337322 PubMed ID: 7612835
TI Interactions of surfactin with membrane models.
AU Maget-Dana R; Ptak M
CS Centre de Biophysique Moleculaire, C.N.R.S., Orleans, France.
SO BIOPHYSICAL JOURNAL, (1995 May) 68 (5) 1937-43.
Journal code: 0370626. ISSN: 0006-3495.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199508
ED Entered STN: 19950905
Last Updated on STN: 19950905
Entered Medline: 19950822
AB Surfactin, an acidic cyclic lipopeptide produced by strains of Bacillus subtilis, is a powerful biosurfactant possessing biological activities. Interactions of ionized surfactin (two **negative** charges) with lecithin vesicles have been monitored by changes in its CD spectra. These changes are more important in the presence of Ca2+ ions. We have studied the penetration of ionized surfactin into lipid monolayers. Using dimyristoyl phospholipids, the surfactin penetration is more important in DMPC than in DMPE monolayers and is greatly reduced in DMPA monolayers because of electrostatic repulsion. The surfactin penetration is lowered when the acyl chain length of the phospholipids increases. The exclusion pressure varies from 40 mN m-1 for DMPC to 30 mN m-1 for DPPC and 18 mN m-1 for egg lecithin. The presence of Ca2+ ions, which neutralize the charges of both surfactin and lipids in the subphase, leads to an important change of the penetration process that is enhanced in the case of acidic, but also of long chain (higher than C14) zwitterionic phospholipids (DPPC and lecithin). From **compression** isotherms of mixed surfactin/phospholipid monolayers, it appears that surfactin is completely miscible with phospholipids. The present study shows that surfactin penetrates spontaneously into lipid membranes by means of hydrophobic interactions. The insertion in the lipid membrane is accompanied by a conformation change of the peptide cycle.

L5 ANSWER 9 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1995:248356 BIOSIS
DN PREV199598262656
TI Hydrolysis of beta-lactoglobulin by thermolysin and pepsin under high hydrostatic pressure.
AU Dufour, Eric; Herve, Guy; Haertle, Tomasz (1)
CS (1) LEIMA, Inst. Natl. Recherche Agronomique, B.P. 527, 44026 Nantes Cedex 03 France
SO Biopolymers, (1995) Vol. 35, No. 5, pp. 475-483.
ISSN: 0006-3525.
DT Article
LA English

AB Hydrolysis of beta-lactoglobulin with thermolysin and pepsin at pressures ranging between 0.1 and 350 MPa showed a significant increase of cleavage rates. Pressure-induced changes of susceptibility to hydrolysis of beta-lactoglobulin proteolytic sites were also observed. The pressure, raised to 200 MPa, accelerates the hydrolysis of beta-lactoglobulin by thermolysin and changes obtained peptide profiles. Initially, higher pressure makes the N-terminal, and to a smaller extent, C-terminal peptide fragments of beta-lactoglobulin molecule, more susceptible to removal by thermolysin. This indicates combined influence of pressure-induced thermolysin activation and partial unfolding of beta-lactoglobulin by **compression** at neutral pHs. The rates of hydrolysis of beta-lactoglobulin by pepsin (negligible at 0.1 MPa) are increased considerably with pressure up to 300 MPa. The susceptibility of beta-lactoglobulin proteolytic sites to peptic cleavage remains constant over all the studied pressure range. The lack of significant qualitative changes in the peptic peptide profiles produced at different pressures and at clearly pressure-dependent rates points to **negative** reaction volume changes as the major factor in peptic hydrolysis of beta-lactoglobulin under high pressure. Thus the beta-lactoglobulin molecule resists pressure-induced unfolding in acid pHs and yields to it in neutral pHs.

L5 ANSWER 10 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1995:156757 BIOSIS
DN PREV199598171057
TI Analysis of errors in finished **DNA sequences**: The surfactin operon of Bacillus subtilis as an example.
AU Fabret, Celine; Quentin, Yves; Guiseppi, Annick; Busuttil, Jeanine; Haiech, Jacques; Denizot, Francois (1)
CS (1) Lab. Chimie Bacterienne, 31 Chemin Joseph Aiguier BP 71, 13277 Marseille Cedex 9 France
SO Microbiology (Reading), (1995) Vol. 141, No. 2, pp. 345-350. ISSN: 1350-0872.
DT Article
LA English
AB Increased productivity in **DNA** sequencing would not be valid without a straightforward detection and estimation of errors in finished **sequences**. The **sequence** of the surfactin operon from Bacillus subtilis was obtained by two different groups and by chance we were also working on the same chromosome region. Taking advantage of this situation we report in this paper, the number and nature of errors found in the overlapping part of the **DNA sequences** obtained by the three laboratories. The coincidence of some of the errors with **compression** in **sequence** ladders and with secondary **DNA** structures as well as the detection of frameshift errors using computer programs, are demonstrated. Finally we discuss the definition of a new sequencing strategy that might minimize both the error rate and the cost of sequencing.

L5 ANSWER 11 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1993:322311 BIOSIS
DN PREV199396030661
TI A-tract and (**positive**)-CC-1065-induced bending of **DNA**: Comparison of structural features using non-denaturing gel analysis, hydroxyl-radical footprinting, and high-field NMR.
AU Sun, Daekyu; Lin, Chin Hsiung; Hurley, Laurence H. (1)
CS (1) Drug Dyn. Inst., Coll. Pharm., Univ. Tex. at Austin, Austin, TX 78712 USA
SO Biochemistry, (1993) Vol. 32, No. 17, pp. 4487-4495. ISSN: 0006-2960.
DT Article
LA English
AB (+)-CC-1065 is a biologically potent **DNA**-reactive antitumor antibiotic produced by Streptomyces zelensis. In a previous study we have reported that (+)-CC-1065 produces bending of **DNA** that has

similarities to that intrinsically associated with A-tracts (Lin, C. H., Sun, D., & Hurley, L. H. (1991) Chem. Res. Toxicol. 4, 21-26). In this article we provide evidence using a combination of non-denaturing gel analysis, hydroxyl-radical footprinting, and high-field NMR for both distinctions between the two types of bends and the importance of junctions in both types of bends. For A-tracts we demonstrate that the locus of bending is at the center of an A-tract and that upon modification of the 3' adenine with (+)-CC-1065 this locus is moved less than 1 base pair to the 3' side, and the bending magnitude is significantly increased. For drug bonding **sequences** such as 5'-AGTTA* or 5'-GATTA* (where * denotes the drug bonding site), the locus of bending is found to be between the two thymines, and the bending is focused over a 2-base-pair **sequence** rather than a 5-base-pair **sequence**, as is the case for the A-tract. An important distinction between an A-tract intrinsic bend and a (+)-CC-1065-induced bend is the effect of temperature. While, as shown previously, the magnitude of A-tract bending increases with decrease in temperature, for drug-induced bending of 5'-AGTTA* the bending magnitude increases with increased temperature. Hydroxyl-radical footprinting of the drug-modified 5'-AGTTA* **sequence** shows a decrease in cleavage centered around the TT **sequence**, which is presumably associated with a decrease in minor groove width. In a parallel study, the non-self-complementary 12-mer duplex (5'-GGCGGAGTTA*GG-3') cntdot (5'-CCTAACTCCGCC-3') (Figure 2B) and the corresponding (+)-CC-1065-modified duplex adduct were examined thoroughly by one- and two-dimensional 1H NMR and NOESY restrained molecular mechanics and dynamics calculations. Both the 12-mer duplex and the (+)-CC-1065-12-mer duplex adduct maintain an overall B-form **DNA** with the anti base orientation throughout in aqueous solution at room temperature. The 18C **nucleotide** of both the 12-mer duplex and its drug-modified adduct has an average C3'-endo sugar pucker. The 12-mer duplex exhibits a unique internal motion at the 16A **nucleotide**, which is located to the 3' side of the complementary partner of the covalently modified adenine, and a major kink at the 18C-19T step. Following covalent bonding with (+)-CC-1065, the discontinuity around 18C is entrapped and further exaggerated. In addition, the 12-mer duplex adduct displays a **compression** of the minor groove at the 8T to 9T step and widening on both sides, but especially abruptly at the covalent modification site. Structurally, the 12-mer duplex adduct bears many similarities to a bent **DNA** structure, which is intrinsically associated with A-tracts. The major drug-induced distortion on **DNA** is localized at the 9T and 10A step of the covalently modified strand. A truncated junction model for the drug-entrapped/induced bending of **DNA** is proposed, and a comparison to intrinsic A-tract bending is made.

L5 ANSWER 12 OF 15 MEDLINE
AN 93256069 MEDLINE
DN 93256069 PubMed ID: 8098180
TI Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism.
AU Lavedan C; Hofmann-Radvanyi H; Shelbourne P; Rabes J P; Duros C; Savoy D; Dehaupas I; Luce S; Johnson K; Junien C
CS INSERM, Unite 73, Paris, France.
SO AMERICAN JOURNAL OF HUMAN GENETICS, (1993 May) 52 (5) 875-83.
Journal code: 0370475. ISSN: 0002-9297.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199306
ED Entered STN: 19930618
Last Updated on STN: 20000303
Entered Medline: 19930607
AB Myotonic dystrophy (DM) is a progressive neuromuscular disorder which results from elongations of an unstable (CTG)n repeat, located in the 3'

untranslated region of the DM gene. A correlation has been demonstrated between the increase in the repeat number of this **sequence** and the severity of the disease. However, the clinical status of patients cannot be unambiguously ascertained solely on the basis of the number of CTG repeats. Moreover, the exclusive maternal inheritance of the congenital form remains unexplained. Our observation of differently sized repeats in various DM tissues from the same individual may explain why the size of the mutation observed in lymphocytes does not necessarily correlate with the severity and nature of symptoms. Through a molecular and genetic study of 142 families including 418 DM patients, we have investigated the dynamics of the CTG repeat meiotic instability. A **positive** correlation between the size of the repeat and the intergenerational enlargement was observed similarly through male and female meioses for < or = 0.5-kb CTG **sequences**. Beyond 0.5 kb, the intergenerational variation was more important through female meioses, whereas a tendency to **compression** was observed almost exclusively in male meioses, for > or = 1.5-kb fragments. This implies a size- and sex-dependent meiotic instability. Moreover, segregation analysis supports the hypothesis of a maternal as well as a familial predisposition for the occurrence of the congenital form. Finally, this analysis reveals a significant excess of transmitting grandfathers partially accounted for by increased fertility in affected males.

L5 ANSWER 13 OF 15 MEDLINE
AN 93368658 MEDLINE
DN 93368658 PubMed ID: 8361538
TI Spondylometaphyseal dysplasia in mice carrying a dominant **negative** mutation in a matrix protein specific for cartilage-to-bone transition.
AU Jacenko O; LuValle P A; Olsen B R
CS Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02115.
SO NATURE, (1993 Sep 2) 365 (6441) 56-61.
Journal code: 0410462. ISSN: 0028-0836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199309
ED Entered STN: 19931015
Last Updated on STN: 19931015
Entered Medline: 19930930
AB The vertebrate skeleton is formed primarily by endochondral ossification, starting during embryogenesis when cartilage anlagens develop central regions of hypertrophic cartilage which are replaced by bony trabeculae and bone marrow. During this process chondrocytes express a unique matrix molecule, type X collagen. We report here that mice carrying a mutated collagen X transgene develop skeletal deformities including **compression** of hypertrophic growth plate cartilage and a decrease in newly formed bone, as well as leukocyte deficiency in bone marrow, reduction in size of thymus and spleen, and lymphopenia. The defects indicate that collagen X is required for normal skeletal morphogenesis and suggest that mutations in COL10A1 are responsible for certain human chondrodysplasias, such as spondylometaphyseal dysplasias and metaphyseal chondrodysplasias.

L5 ANSWER 14 OF 15 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1989:123948 BIOSIS
DN BA87:58601
TI **DNA** SEQUENCING WITH THERMUS-AQUATICUS **DNA** POLYMERASE AND DIRECT SEQUENCING OF POLYMERASE CHAIN REACTION-AMPLIFIED **DNA**
.
AU INNIS M A; MYAMBO K B; GELFAND D H; BROW M A D
CS DEP. MICROBIAL GENETICS, CETUS CORP., 1400 FIFTY-THIRD ST., EMERYVILLE, CA 94608.
SO PROC NATL ACAD SCI U S A, (1988) 85 (24), 9436-9440.

CODEN: PNASA6. ISSN: 0027-8424.
FS BA; OLD
LA English
AB The highly thermostable **DNA** polymerase from Thermus aquaticus (Taq) is ideal for both manual and automated **DNA** sequencing because it is fast, highly processive, has little or no 3'-exonuclease activity, and is active over a broad range of temperatures. Sequencing protocols are presented that produce readable extension products > 1000 bases having uniform band intensities. A combination of high reaction temperatures and the base analog 7-deaza-2'-deoxyguanosine was used to **sequence** through G+C-rich **DNA** and to resolve gel **compressions**. We modified the polymerase chain reaction (PCR) conditions for direct **DNA** sequencing of asymmetric PCR products without intermediate purification by using Taq **DNA** polymerase. The coupling of template preparation by asymmetric PCR and direct sequencing should facilitate automation for large-scale sequencing projects.

L5 ANSWER 15 OF 15 MEDLINE
AN 90148261 MEDLINE
DN 90148261 PubMed ID: 6400899
TI A stochastic model for helix bending in B-**DNA**.
AU Dickerson R E; Kopka M L; Pjura P
CS Molecular Biology Institute, University of California, Los Angeles 90024.
NC GM-30543 (NIGMS)
SO JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS, (1983 Dec) 1 (3) 755-71.
Journal code: 8404176. ISSN: 0739-1102.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199003
ED Entered STN: 19900601
Last Updated on STN: 19970203
Entered Medline: 19900326
AB Bending in double-helical B-**DNA** apparently occurs only by rolling adjacent base pairs over one another along their long axes. The lifting apart of ends that would be required by tilt or wedge angle contributions is too costly in free energy and does not occur. Roll angles at base steps can be **positive** (**compression** of major groove) or **negative** (**compression** of minor groove), with the former somewhat easier. Individual steps may advance or oppose the overall direction of bend, or make lateral excursions, but the result of this series of "random roll" steps is the production of a net bending in the helix axis. Because the natural roll points for bending in a given plane occur every 5 base pairs, one would expect that double-helical **DNA** wrapped around a nucleosome core would exhibit bends with the same periodicity. Alternate bends might be particularly acute where the major groove faced the nucleosome core and was compressed against it. The "annealed kinking" model proposed by Fratini et al. (J. Biol. Chem. 257, 14686 (1982) was suggested from the observation that a major bend at a natural roll point is flanked by decreasing roll angles at the steps to either side, as though local strain was being minimized by somewhat blurring the bend out rather than keeping it localized. The random walk model suggested in this paper would describe this as a decreased roll angle as the helix step rotates toward a direction perpendicular to the overall bend. Bending of **DNA** is seen to be a more stochastic process than had been suspected. Detailed analysis of every helix step reveals both side excursions and backward or retrograde motion, as in any random walk situation. Yet these isolated steps counteract one another, to leave behind a residuum of overall bending in a specific direction.

=>

---Logging off of STN---

=>
Executing the logoff script...

=> LOG Y

COST IN U.S. DOLLARS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	17.81	18.02

STN INTERNATIONAL LOGOFF AT 17:53:06 ON 04 SEP 2002

Search Paper 6

Welcome to STN International! Enter x:x

LOGINID:SSSPTA1805JXB

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2

* * * * * * * * * * Welcome to STN International * * * * * * * * * *

NEWS 1		Web Page URLs for STN Seminar Schedule - N. America
NEWS 2	Apr 08	"Ask CAS" for self-help around the clock
NEWS 3	Apr 09	BEILSTEIN: Reload and Implementation of a New Subject Area
NEWS 4	Apr 09	ZDB will be removed from STN
NEWS 5	Apr 19	US Patent Applications available in IFICDB, IFIPAT, and IFIUDB
NEWS 6	Apr 22	Records from IP.com available in CAPLUS, HCAPLUS, and ZCAPLUS
NEWS 7	Apr 22	BIOSIS Gene Names now available in TOXCENTER
NEWS 8	Apr 22	Federal Research in Progress (FEDRIP) now available
NEWS 9	Jun 03	New e-mail delivery for search results now available
NEWS 10	Jun 10	MEDLINE Reload
NEWS 11	Jun 10	PCTFULL has been reloaded
NEWS 12	Jul 02	FOREGE no longer contains STANDARDS file segment
NEWS 13	Jul 22	USAN to be reloaded July 28, 2002; saved answer sets no longer valid
NEWS 14	Jul 29	Enhanced polymer searching in REGISTRY
NEWS 15	Jul 30	NETFIRST to be removed from STN
NEWS 16	Aug 08	CANCERLIT reload
NEWS 17	Aug 08	PHARMAMarketLetter(PHARMAML) - new on STN
NEWS 18	Aug 08	NTIS has been reloaded and enhanced
NEWS 19	Aug 19	Aquatic Toxicity Information Retrieval (AQUIRE) now available on STN
NEWS 20	Aug 19	IFIPAT, IFICDB, and IFIUDB have been reloaded
NEWS 21	Aug 19	The MEDLINE file segment of TOXCENTER has been reloaded
NEWS 22	Aug 26	Sequence searching in REGISTRY enhanced
NEWS 23	Sep 03	JAPIO has been reloaded and enhanced

NEWS EXPRESS	February 1 CURRENT WINDOWS VERSION IS V6.0d, CURRENT MACINTOSH VERSION IS V6.0a(ENG) AND V6.0Ja(JP), AND CURRENT DISCOVER FILE IS DATED 05 FEBRUARY 2002
NEWS HOURS	STN Operating Hours Plus Help Desk Availability
NEWS INTER	General Internet Information
NEWS LOGIN	Welcome Banner and News Items
NEWS PHONE	Direct Dial and Telecommunication Network Access to STN
NEWS WWW	CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that specific topic.

All use of STN is subject to the provisions of the STN Customer agreement. Please note that this agreement limits use to scientific research. Use for software development or design or implementation of commercial gateways or other similar uses is prohibited and may result in loss of user privileges and other penalties.

* * * * * * * * * * * * * * STN Columbus * * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 18:18:03 ON 04 SEP 2002

=> file .pub
COST IN U.S. DOLLARS SINCE FILE TOTAL

```
                                                        ENTRY    SESSION
FULL ESTIMATED COST                                      0.21       0.21

FILE 'MEDLINE' ENTERED AT 18:18:14 ON 04 SEP 2002

FILE 'BIOSIS' ENTERED AT 18:18:14 ON 04 SEP 2002
COPYRIGHT (C) 2002 BIOLOGICAL ABSTRACTS INC.(R)

=> s digital signal processor
L1          122 DIGITAL SIGNAL PROCESSOR

=> s l1 and sequence
L2            8 L1 AND SEQUENCE

=> duplicate remove l2
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE?  Y/(N):n
PROCESSING COMPLETED FOR L2
L3             5 DUPLICATE REMOVE L2 (3 DUPLICATES REMOVED)

=> d 1-5 bib ab

L3   ANSWER 1 OF 5     MEDLINE                          DUPLICATE 1
AN   2001122976     MEDLINE
DN   21017623   PubMed ID: 11144586
TI   A concept for a research tool for experiments with cochlear implant users.
AU   Geurts L; Wouters J
CS   Laboratoire Experimental ORL, KULeuven, Belgium..
     Luc.Geurts@uz.kuleuven.ac.be
SO   JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, (2000 Dec) 108 (6) 2949-56.
     Journal code: 7503051. ISSN: 0001-4966.
CY   United States
DT   Journal; Article; (JOURNAL ARTICLE)
LA   English
FS   Priority Journals
EM   200102
ED   Entered STN: 20010322
     Last Updated on STN: 20010322
     Entered Medline: 20010222
AB   APEX, an acronym for computer Application for Psycho-Electrical
     eXperiments, is a user friendly tool used to conduct psychophysical
     experiments and to investigate new speech coding algorithms with cochlear
     implant users. Most common psychophysical experiments can be easily
     programmed and all stimuli can be easily created without any knowledge of
     computer programing. The pulsatile stimuli are composed off-line using
     custom-made MATLAB (Registered trademark of The Mathworks, Inc.,
     http://www.mathworks.com) functions and are stored on hard disk or CD ROM.
     These functions convert either a speech signal into a pulse
     sequence or generate any sequence of pulses based on the
     parameters specified by the experimenter. The APEX personal computer (PC)
     software reads a text file which specifies the experiment and the stimuli,
     controls the experiment, delivers the stimuli to the subject through a
     digital signal processor (DSP) board, collects
     the responses via a computer mouse or a graphics tablet, and writes the
     results to the same file. At present, the APEX system is implemented for
     the LAURA (Registered trademark of Philips Hearing Implants) cochlear
     implant. However, the concept-and many parts of the system-is portable to
     any other device. Also, psycho-acoustical experiments can be conducted by
     presenting the stimuli acoustically through a sound card.

L3   ANSWER 2 OF 5  BIOSIS  COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN   2001:307072  BIOSIS
DN   PREV200100307072
TI   Dynamic infrared imaging (DIRI) of newly diagnosed lymphoma. Correlation
     with gallium-67 imaging (GA) and F-18 FDG positron emission tomography
```

(PET.
AU Janicek, Milos J. (1); Janicek, Milos R. (1); Friedberg, Jonathan W. (1); Demetri, George D. (1)
CS (1) Radiology and Adult Oncology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA USA
SO Blood, (November 16, 2000) Vol. 96, No. 11 Part 2, pp. 224b-225b. print. Meeting Info.: 42nd Annual Meeting of the American Society of Hematology San Francisco, California, USA December 01-05, 2000 American Society of Hematology
. ISSN: 0006-4971.
DT Conference
LA English
SL English
AB Staging of malignant lymphoma relies heavily on conventional imaging, combining measurement of masses on computed tomography (CT) with radionuclide-based functional studies, which are relatively non-specific. Treatment decisions are based on early assessment of response to therapy and residual disease. DIRI may add new dimension to functional non-invasive imaging recording in digitized in plane manner oscillations of temperature and heat distribution in tumors as well as normal tissues. CT, Ga-67, PET in nine patients (7 male, 2 female) with newly diagnosed malignant lymphoma (5 H.D., 4 NHL) affecting neck, axilae, and anterior mediastinum were compared at the time of staging with dynamic infrared images of tumors in selected superficial locations. 28 tumor sites were identified on GA and PET with CT measurable tumor masses amenable to single view dynamic acquisition utilizing BioScan System (OmniCorder Technologies Inc., Stony Brook, N.Y.) equipped with a 256X256 quantum well infrared photodetector (QWIP) focal plan array (FPA) taking **sequence** of 2048 images over 20 sec processed by 32-bit **digital signal processor**. Heat map changes in 8 to 10 micron range characteristic for body temperature were analyzed for raw temperature profile, modulation of temperature, and homogeneity of heat distribution in all 28 tumor sites and compared to 29 regions in adjacent soft tissues with no detectable disease. Very close correlations were observed between tumor depiction by GA, PET, CT, and DIRI. Average temperature sampled over tumor masses with 20X20 pixel region was significantly higher that over adjacent soft tissues (31.79 +/-0.98 vs. 31.23 +/-0.71 dgr C: P=0.17, t-test) confirmed on color-coded maps as areas of high relative temperature and high temperature modulation. On semi-quantitative evaluation, there was significant correlation between GA uptake and high temperature modulation (P<.001, Fisher Exact test), as well as relative temperature assessment (P<.001) comparing tumor sites with soft tissues outside tumor. No consistent pattern of homogeneity of temperature distribution throughout tumors was seen compared to normal tissues (P=.375, Chi-square). In conclusion, raw temperature and temperature modulation measured with DIRI is able to distinguish lymphoma from adjacent tissues. Prospective studies are underway incorporating this promising functional imaging technique to follow patients with lymphoma after therapy.

L3 ANSWER 3 OF 5 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1997:448668 BIOSIS
DN PREV199799747871
TI Advanced multiplexed analysis with the FlowMetrix system.
AU Fulton, R. Jerrold (1); McDade, Ralph L.; Smith, Perry L.; Kienker, Laura J.; Kettman, John R., Jr.
CS (1) Luminex Corp., 1638 Osprey Dr., DeSoto, TX 75115 USA
SO Clinical Chemistry, (1997) Vol. 43, No. 9, pp. 1749-1756.
ISSN: 0009-9147.
DT Article
LA English
AB The FlowMetrix System is a multiplexed data acquisition and analysis platform for flow cytometric analysis of microsphere-based assays that performs simultaneous measurement of up to 64 different analytes. The system consists of 64 distinct sets of fluorescent microspheres and a

standard benchtop flow cytometer interfaced with a personal computer containing a digital signal processing board and Windows95-based software. Individual sets of microspheres can be modified with reactive components such as antigens, antibodies, or oligonucleotides, and then mixed to form a multiplexed assay set. The digital signal-processing hardware and Windows95-based software provide complete control of the flow cytometer and perform real-time data processing, allowing multiple independent reactions to be analyzed simultaneously. The system has been used to perform qualitative and quantitative immunoassays for multiple serum proteins in both capture and competitive inhibition assay formats. The system has also been used to perform DNA **sequence** analysis by multiplexed competitive hybridization with 16 different **sequence** -specific oligonucleotide probes.

L3 ANSWER 4 OF 5 MEDLINE DUPLICATE 2
AN 96105579 MEDLINE
DN 96105579 PubMed ID: 8529353
TI An instrument for real-time spectral estimation of heart rate variability signals.
AU Basano L; Ottonello P; Poggi A; Adezati L; Semino S; Ubaldi P; Viviani G L
CS Dipartimento di Fisica, Universita di Genova, Italy.
SO COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, (1995 Aug) 47 (3) 229-36.
Journal code: 8506513. ISSN: 0169-2607.
CY Ireland
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199601
ED Entered STN: 19960220
Last Updated on STN: 19960220
Entered Medline: 19960130
AB A **Digital Signal Processor** (DSP)-based instrument is proposed for estimating and displaying the Heart Rate Variability (HRV) spectrum in real-time. It consists of an intelligent module which is properly interfaced to an IBM PC and whose operations are independent from the computer's other tasks. In this way, the simultaneous recording of the ECG **sequence**, needed for the more complete off-line analysis, can be performed by the same host. The employed hybrid spectral estimator (in which a classical FFT analysis follows the autoregressive extrapolation of data) appears to be the most apt for the present fixed point arithmetics implementation. The reliability of the instrument and its accuracy are checked both with suitable test signals and by comparison with the results obtained through off-line analysis of the same ECG tracks. The instrument is presently used for cardiovascular investigations, in particular for quickly picking patients with cardiac autonomic neuropathy (CAN) out of a population of diabetic subjects.

L3 ANSWER 5 OF 5 MEDLINE DUPLICATE 3
AN 89090730 MEDLINE
DN 89090730 PubMed ID: 3208622
TI Imaging system for morphometric assessment of absorption or fluorescence in stained cells.
AU Jaggi B; Poon S S; MacAulay C; Palcic B
CS Cancer Imaging, B.C. Cancer Research Centre, Vancouver, British Columbia, Canada.
SO CYTOMETRY, (1988 Nov) 9 (6) 566-72.
Journal code: 8102328. ISSN: 0196-4763.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 198902
ED Entered STN: 19900308
Last Updated on STN: 19900308
Entered Medline: 19890223

AB An image acquisition and processing system has been developed for quantitative microscopy of absorption or fluorescence in stained cells. Three different light transducers are used in the system to exploit the best characteristics of these sensors for different biological measurements. A digital scanner, in the form of a linear array charge-coupled device (CCD), acquires data with high spatial and photometric resolution. A color (RGB) camera is employed when spectral information is required for the segmentation of cellular subcomponents. An image-intensified charged-injection device (CID) camera provides for very low light intensity measurements, primarily for fluorescence-labeled cells. Properties of these transducers, such as contrast transfer function, linearity, and photo-response nonuniformity, have been measured. Two dedicated image processing units were incorporated into the system. The front-end processor, based on a **digital signal processor**, provides functions such as object detection, raw image calibration, compression, artifact removal, and filtering. The second image processor is associated with the frame memory and includes a histogram processor, a dedicated arithmetic logic unit for image processing functions, and a graphics module for one-bit overlay functions. An interactive program was developed to acquire cell images and to experiment with a range of segmentation algorithms, feature extractions, and other image processing functions. The results of any image operation are displayed on the video monitor. Once a desired processing **sequence** is determined, the **sequence** may be stored to become part of a command library and can be executed thereafter as a single instruction.

```
=> d his

     (FILE 'HOME' ENTERED AT 18:18:03 ON 04 SEP 2002)

     FILE 'MEDLINE, BIOSIS' ENTERED AT 18:18:14 ON 04 SEP 2002
L1           122 S DIGITAL SIGNAL PROCESSOR
L2             8 S L1 AND SEQUENCE
L3             5 DUPLICATE REMOVE L2 (3 DUPLICATES REMOVED)

=> s digital signal process?
L4           553 DIGITAL SIGNAL PROCESS?

=> s l4 and sequence
L5            21 L4 AND SEQUENCE

=> s l5 not l2
L6            13 L5 NOT L2

=> duplicate remove l6
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE?  Y/(N):n
PROCESSING COMPLETED FOR L6
L7             9 DUPLICATE REMOVE L6 (4 DUPLICATES REMOVED)

=> d 1-9 bib ab

L7   ANSWER 1 OF 9     MEDLINE                              DUPLICATE 1
AN   2001212851     MEDLINE
DN   21111527   PubMed ID: 11159326
TI   Frequency-domain analysis of biomolecular sequences.
AU   Anastassiou D
CS   Department of Electrical Engineering, Columbia University, 500 West 120th
     Street, Mail Code 4712, New York, NY 10027, USA.. anastas@ee.columbia.edu
SO   BIOINFORMATICS, (2000 Dec) 16 (12) 1073-81.
     Journal code: 9808944. ISSN: 1367-4803.
CY   England: United Kingdom
DT   Journal; Article; (JOURNAL ARTICLE)
```

LA English
FS Priority Journals
EM 200104
ED Entered STN: 20010425
Last Updated on STN: 20010425
Entered Medline: 20010419
AB MOTIVATION: Frequency-domain analysis of biomolecular **sequences** is hindered by their representation as strings of characters. If numerical values are assigned to each of these characters, then the resulting numerical **sequences** are readily amenable to **digital signal processing**. RESULTS: We introduce new computational and visual tools for biomolecular **sequences** analysis. In particular, we provide an optimization procedure improving upon traditional Fourier analysis performance in distinguishing coding from noncoding regions in DNA **sequences**. We also show that the phase of a properly defined Fourier transform is a powerful predictor of the reading frame of protein coding regions. Resulting color maps help in visually identifying not only the existence of protein coding areas for both DNA strands, but also the coding direction and the reading frame for each of the exons. Furthermore, we demonstrate that color spectrograms can visually provide, in the form of local 'texture', significant information about biomolecular **sequences**, thus facilitating understanding of local nature, structure and function.

L7 ANSWER 2 OF 9 MEDLINE DUPLICATE 2
AN 1999178287 MEDLINE
DN 99178287 PubMed ID: 10080202
TI Very high-frequency ultrasound corneal analysis identifies anatomic correlates of optical complications of lamellar refractive surgery: anatomic diagnosis in lamellar surgery.
AU Reinstein D Z; Silverman R H; Sutton H F; Coleman D J
CS Department of Ophthalmology, Weill College of Medicine of Cornell University, New York, New York 10021, USA.
NC EY01212 (NEI)
SO OPHTHALMOLOGY, (1999 Mar) 106 (3) 474-82.
Journal code: 7802443. ISSN: 0161-6420.
CY United States
DT (CLINICAL TRIAL)
Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199903
ED Entered STN: 19990326
Last Updated on STN: 19990326
Entered Medline: 19990318
AB OBJECTIVE: To examine the utility of very high-frequency (VHF) ultrasound scanning in determining the anatomic changes and correlates of optical complications in lamellar refractive surgery. STUDY DESIGN: Case series. PARTICIPANTS: Cases analyzed included marked asymmetric astigmatism postautomated lamellar keratoplasty (ALK), image ghosting despite normal videokeratography post-ALK, uncomplicated myopic laser in situ keratomileusis (LASIK), and hyperopic LASIK with regression. METHODS: A prototype VHF ultrasound scanner (50 MHz) was used to obtain **sequences** of parallel B-scans of the cornea. **Digital signal processing** techniques were used to measure epithelial, stromal, and flap thickness values in a grid encompassing the central 4 to 5 mm of the cornea, enabling pachymetric mapping of each layer with 2-micron precision. MAIN OUTCOME MEASURE: The appearance of the corneas in VHF ultrasound images and thickness values of individual corneal layers determined from VHF ultrasound data. RESULTS: VHF ultrasound resolved the epithelial, stromal cap, or flap and residual stromal layers 1 year after lamellar surgery. Asymmetric stromal tissue removal was differentiated from stromal cap irregularity. Epithelium acted to compensate for asymmetry of the stromal surface about the visual axis and for localized surface irregularities. Irregularities in the

epithelial-stromal interface accounted for image ghosting present despite apparently normal videokeratography. Epithelial thickening was shown after uncomplicated myopic LASIK. Hyperopic LASIK demonstrated relative epithelial thickening localized to the region of ablation accounting for refractive regression. CONCLUSIONS: VHF ultrasound shows promise as a sensitive method of determining the anatomic correlates of optical complications in lamellar refractive surgery.

L7 ANSWER 3 OF 9 MEDLINE
AN 97444982 MEDLINE
DN 97444982 PubMed ID: 9299971
TI Advanced multiplexed analysis with the FlowMetrix system.
AU Fulton R J; McDade R L; Smith P L; Kienker L J; Kettman J R Jr
CS Luminex Corporation, Austin, TX 78727, USA.. jfulton@luminexcorp.com
SO CLINICAL CHEMISTRY, (1997 Sep) 43 (9) 1749-56.
Journal code: 9421549. ISSN: 0009-9147.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199710
ED Entered STN: 19971021
Last Updated on STN: 19971021
Entered Medline: 19971009
AB The FlowMetrix System is a multiplexed data acquisition and analysis platform for flow cytometric analysis of microsphere-based assays that performs simultaneous measurement of up to 64 different analytes. The system consists of 64 distinct sets of fluorescent microspheres and a standard benchtop flow cytometer interfaced with a personal computer containing a **digital signal processing** board and Windows95-based software. Individual sets of microspheres can be modified with reactive components such as antigens, antibodies, or oligonucleotides, and then mixed to form a multiplexed assay set. The **digital signal-processing** hardware and Windows95-based software provide complete control of the flow cytometer and perform real-time data processing, allowing multiple independent reactions to be analyzed simultaneously. The system has been used to perform qualitative and quantitative immunoassays for multiple serum proteins in both capture and competitive inhibition assay formats. The system has also been used to perform DNA **sequence** analysis by multiplexed competitive hybridization with 16 different **sequence** -specific oligonucleotide probes.

L7 ANSWER 4 OF 9 MEDLINE
AN 1998196224 MEDLINE
DN 98196224 PubMed ID: 9534725
TI New and future developments in ultrasonic imaging.
AU Whittingham T A
CS Regional Medical Physics Department, Newcastle General Hospital, Newcastle upon Tyne, UK.
SO BRITISH JOURNAL OF RADIOLOGY, (1997 Nov) 70 Spec No S119-32. Ref: 11
Journal code: 0373125. ISSN: 0007-1285.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Abridged Index Medicus Journals; Priority Journals
EM 199804
ED Entered STN: 19980430
Last Updated on STN: 19980430
Entered Medline: 19980421
AB In the first part of the review, recent developments in medical imaging technology are described. Developments in transducer materials and matching, leading to improvements in band-width and sensitivity are

discussed. Improvements in dynamic range due to increased transducer sensitivity, lower electronic noise levels and more efficient filtering are then considered. The benefits of the application of **digital signal processing** (DSP) techniques to radiofrequency (RF) echo signals are described, including more precise filtering and beam forming, synthetic aperture and parallel receive beam forming. Finally, the current situation in regard to 1.5 D arrays, 3 D scanning, ultrasound computed tomography (UCT), harmonic imaging with contrast agents and elastography are discussed. In the second part, some predictions for future developments are made. These will be possible largely due to the power of DSP. Parallel transmissions will make more efficient use of time, allowing greater spatial and temporal resolution, and greater accuracy in Doppler imaging. Adaptive transmission tailoring will be used, where the pulse characteristics to each part of the image field are independently optimized, as will adaptive receive processing in which echo **sequences** from each part of the image are independently and optimally processed. An important potential development will be automatic feature recognition, making possible accurate compound scanning with high spatial resolution, and quantitative information about the spatial distribution of acoustic speed. Compound scanning will provide more complete visualization of all structures and, particularly when incorporated into intravascular probes, should greatly aid the investigation of arterial plaque morphology. Feature recognition will also make it possible to have UCT systems (array based in future) which require less than 360 degrees access. Harmonic imaging without contrast agents, based simply on the inherent non-linearity of sound propagation in tissue, will become common. 2 D phased array transducer will permit symmetric beam focusing and scanning throughout a solid cone, greatly facilitating the development of 3 D scanning applications. Large 2 D arrays would have the potential to produce a five-fold increase in spatial resolution of a limited volume of tissue, or to measure the variation of backscatter with angle, as an aid to tissue characterization. Finally, ultrasound will be increasingly used to measure the elastic and dynamic properties of local regions of tissue.

L7 ANSWER 5 OF 9 MEDLINE
AN 97021030 MEDLINE
DN 97021030 PubMed ID: 8867390
TI Preliminary expansion of the resonant recognition model to incorporate multi variable analysis.
AU Birch S; West R; Cosic I
CS Department of Electrical and Computer Systems Engineering, Monash University, Caulfield, Victoria.
SO AUSTRALASIAN PHYSICAL AND ENGINEERING SCIENCES IN MEDICINE, (1995 Dec) 18 (4) 197-207.
Journal code: 8208130. ISSN: 0158-9938.
CY Australia
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199612
ED Entered STN: 19970128
Last Updated on STN: 19970128
Entered Medline: 19961210
AB The Resonant Recognition Model (rrm) uses **digital signal processing** methods to investigate protein structure-function; and links the biological function of protein families to unique characteristic frequencies. The rrm originally used a single set of variables: the electron ion interaction potential (EIIP). Here the rrm has been expanded to include 242 sets of variables to analyse a sample of protein families. Despite the evident increase in complexity of the data, distinguishing patterns can be observed between the different protein families. The thus-obtained Signature Profiles (SP) indicate that proteins having similar overall functions may be identifiable and differentiated from others by their characteristic frequency signatures far more readily than

with the single variable rrm spectra.

L7 ANSWER 6 OF 9 MEDLINE DUPLICATE 3
AN 95154596 MEDLINE
DN 95154596 PubMed ID: 7851673
TI Correlation of cervical auscultation with physiological recording during suckle-feeding in newborn infants.
AU Vice F L; Bamford O; Heinz J M; Bosma J F
CS Department of Pediatrics, University of Maryland Hospital, Baltimore 21201.
SO DEVELOPMENTAL MEDICINE AND CHILD NEUROLOGY, (1995 Feb) 37 (2) 167-79. Journal code: 0006761. ISSN: 0012-1622.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199503
ED Entered STN: 19950322
Last Updated on STN: 19950322
Entered Medline: 19950316
AB Pharyngeal swallows during infant suckle-feeding are associated with a characteristic **sequence** of sounds audible by stethoscope or by an accelerometer or microphone held over the larynx. In rhythmically feeding term-born neonates, the delineating acoustic elements are discrete sounds which precede and succeed pharyngeal swallows. **Digital signal processing** shows similarities in morphological detail between the discrete sounds preceding swallows and between those succeeding swallows; those succeeding swallows are more variable in temporal relation to swallows, amplitude and morphological detail. Variations in the pattern of interswallow respiration, including apnea, are correlated with variations in the discrete sounds. Specification of physiological correlates of these internal feeding sounds increases the utility of cervical auscultation as a method of investigation and of clinical observation of feeding.

L7 ANSWER 7 OF 9 MEDLINE DUPLICATE 4
AN 90272406 MEDLINE
DN 90272406 PubMed ID: 2349096
TI **Digital signal processing** methods for biosequence comparison.
AU Benson D C
CS Department of Mathematics, University of California, Davis 95616.
SO NUCLEIC ACIDS RESEARCH, (1990 May 25) 18 (10) 3001-6. Journal code: 0411011. ISSN: 0305-1048.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199007
ED Entered STN: 19900810
Last Updated on STN: 19900810
Entered Medline: 19900711
AB A method is discussed for DNA or protein **sequence** comparison using a finite field fast Fourier transform, a **digital signal processing** technique; and statistical methods are discussed for analyzing the output of this algorithm. This method compares two **sequences** of length N in computing time proportional to N log N compared to N2 for methods currently used. This method makes it feasible to compare very long **sequences**. An example is given to show that the method correctly identifies sites of known homology.

L7 ANSWER 8 OF 9 MEDLINE
AN 91095465 MEDLINE
DN 91095465 PubMed ID: 1702543
TI Characterization of single channel currents using **digital**

signal processing techniques based on Hidden Markov Models.
AU Chung S H; Moore J B; Xia L G; Premkumar L S; Gage P W
CS Research School of Biological Sciences, Australian National University, Canberra.
SO PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON. SERIES B: BIOLOGICAL SCIENCES, (1990 Sep 29) 329 (1254) 265-85.
Journal code: 7503623. ISSN: 0962-8436.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199102
ED Entered STN: 19910322
Last Updated on STN: 19990129
Entered Medline: 19910211
AB Techniques for extracting small, single channel ion currents from background noise are described and tested. It is assumed that single channel currents are generated by a first-order, finite-state, discrete-time, Markov process to which is added 'white' background noise from the recording apparatus (electrode, amplifiers, etc). Given the observations and the statistics of the background noise, the techniques described here yield a posteriori estimates of the most likely signal statistics, including the Markov model state transition probabilities, duration (open- and closed-time) probabilities, histograms, signal levels, and the most likely state **sequence**. Using variations of several algorithms previously developed for solving digital estimation problems, we have demonstrated that: (1) artificial, small, first-order, finite-state, Markov model signals embedded in simulated noise can be extracted with a high degree of accuracy, (2) processing can detect signals that do not conform to a first-order Markov model but the method is less accurate when the background noise is not white, and (3) the techniques can be used to extract from the baseline noise single channel currents in neuronal membranes. Some studies have been included to test the validity of assuming a first-order Markov model for biological signals. This method can be used to obtain directly from digitized data, channel characteristics such as amplitude distributions, transition matrices and open- and closed-time durations.

L7 ANSWER 9 OF 9 MEDLINE
AN 85205977 MEDLINE
DN 85205977 PubMed ID: 2581884
TI Is it possible to analyze DNA and protein **sequences** by the methods of **digital signal processing**?.
AU Veljkovic V; Cosic I; Dimitrijevic B; Lalovic D
SO IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, (1985 May) 32 (5) 337-41.
Journal code: 0012737. ISSN: 0018-9294.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 198507
ED Entered STN: 19900320
Last Updated on STN: 19900320
Entered Medline: 19850725

=>

---Logging off of STN---

=>
Executing the logoff script...

=> LOG Y

COST IN U.S. DOLLARS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	12.40	12.61

STN INTERNATIONAL LOGOFF AT 18:24:34 ON 04 SEP 2002

Search Paper 8

Welcome to STN International! Enter x:x

LOGINID:SSSPTA1805JXB

PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2

* * * * * * * * * * Welcome to STN International * * * * * * * * * *

NEWS 1		Web Page URLs for STN Seminar Schedule - N. America
NEWS 2	Apr 08	"Ask CAS" for self-help around the clock
NEWS 3	Apr 09	BEILSTEIN: Reload and Implementation of a New Subject Area
NEWS 4	Apr 09	ZDB will be removed from STN
NEWS 5	Apr 19	US Patent Applications available in IFICDB, IFIPAT, and IFIUDB
NEWS 6	Apr 22	Records from IP.com available in CAPLUS, HCAPLUS, and ZCAPLUS
NEWS 7	Apr 22	BIOSIS Gene Names now available in TOXCENTER
NEWS 8	Apr 22	Federal Research in Progress (FEDRIP) now available
NEWS 9	Jun 03	New e-mail delivery for search results now available
NEWS 10	Jun 10	MEDLINE Reload
NEWS 11	Jun 10	PCTFULL has been reloaded
NEWS 12	Jul 02	FOREGE no longer contains STANDARDS file segment
NEWS 13	Jul 22	USAN to be reloaded July 28, 2002; saved answer sets no longer valid
NEWS 14	Jul 29	Enhanced polymer searching in REGISTRY
NEWS 15	Jul 30	NETFIRST to be removed from STN
NEWS 16	Aug 08	CANCERLIT reload
NEWS 17	Aug 08	PHARMAMarketLetter(PHARMAML) - new on STN
NEWS 18	Aug 08	NTIS has been reloaded and enhanced
NEWS 19	Aug 19	Aquatic Toxicity Information Retrieval (AQUIRE) now available on STN
NEWS 20	Aug 19	IFIPAT, IFICDB, and IFIUDB have been reloaded
NEWS 21	Aug 19	The MEDLINE file segment of TOXCENTER has been reloaded
NEWS 22	Aug 26	Sequence searching in REGISTRY enhanced
NEWS 23	Sep 03	JAPIO has been reloaded and enhanced

NEWS EXPRESS	February 1 CURRENT WINDOWS VERSION IS V6.0d, CURRENT MACINTOSH VERSION IS V6.0a(ENG) AND V6.0Ja(JP), AND CURRENT DISCOVER FILE IS DATED 05 FEBRUARY 2002
NEWS HOURS	STN Operating Hours Plus Help Desk Availability
NEWS INTER	General Internet Information
NEWS LOGIN	Welcome Banner and News Items
NEWS PHONE	Direct Dial and Telecommunication Network Access to STN
NEWS WWW	CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that specific topic.

All use of STN is subject to the provisions of the STN Customer agreement. Please note that this agreement limits use to scientific research. Use for software development or design or implementation of commercial gateways or other similar uses is prohibited and may result in loss of user privileges and other penalties.

* * * * * * * * * * * * * * * * STN Columbus * * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 17:01:59 ON 04 SEP 2002

=> file .pub
COST IN U.S. DOLLARS SINCE FILE TOTAL

	ENTRY	SESSION
FULL ESTIMATED COST	0.21	0.21

FILE 'MEDLINE' ENTERED AT 17:02:09 ON 04 SEP 2002

FILE 'BIOSIS' ENTERED AT 17:02:09 ON 04 SEP 2002

=> sequence and bits
SEQUENCE IS NOT A RECOGNIZED COMMAND
The previous command name entered was not recognized by the system.
For a list of commands available to you in the current file, enter
"HELP COMMANDS" at an arrow prompt (=>).

=> s sequence and bits
L1 128 SEQUENCE AND BITS

=> d 1-5 bib ab

L1 ANSWER 1 OF 128 MEDLINE
AN 2002434995 IN-PROCESS
DN 22180120 PubMed ID: 12190442
TI Quantum Information is Incompressible Without Errors.
AU Koashi Masato; Imoto Nobuyuki
CS CREST Research Team for Interacting Carrier Electronics, School of Advanced Sciences, The Graduate University for Advanced Studies (SOKEN), Hayama, Kanagawa, 240-0193, Japan.
SO PHYSICAL REVIEW LETTERS, (2002 Aug 26) 89 (9) 097904.
Journal code: 0401141. ISSN: 0031-9007.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS IN-PROCESS; NONINDEXED; Priority Journals
ED Entered STN: 20020823
Last Updated on STN: 20020823
AB A classical random variable can be faithfully compressed into a **sequence** of **bits** with its expected length lying within one bit of Shannon entropy. We generalize this variable-length and faithful scenario to the general quantum source producing mixed states rho(i) with probability p(i). In contrast to the classical case, the optimal compression rate in the limit of large block length differs from the one in the fixed-length and asymptotically faithful scenario. The amount of this gap is interpreted as the genuinely quantum part being incompressible in the former scenario.

L1 ANSWER 2 OF 128 MEDLINE
AN 2002327328 IN-PROCESS
DN 22065348 PubMed ID: 12070753
TI Coding of disparity information in extrastriate cortex of the cat.
AU Vickery R M; Morley J W
CS School of Physiology and Pharmacology, University of New South Wales, Sydney 2052, Australia,. richard.vickery@unsw.edu.au
SO EXPERIMENTAL BRAIN RESEARCH, (2002 Jul) 145 (1) 130-2.
Journal code: 0043312. ISSN: 0014-4819.
CY Germany: Germany, Federal Republic of
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS IN-PROCESS; NONINDEXED; Priority Journals
ED Entered STN: 20020619
Last Updated on STN: 20020619
AB We have used information theory to analyse the responses of neurons in area 21a of the cat to disparity stimuli. Visual stimuli consisted of drifting sinusoidal gratings presented simultaneously to each eye. The relative spatial phase of the gratings varied between stimulus periods in a pseudo-random **sequence** of 45 degrees increments that covered

the full 360 degrees. The mean information content of the responses of all neurons across all phases was 0.72 **bits** (+/-0.10, SE, n=29). The information conveyed by each neuron was well correlated with the extent to which the interocular phase difference modulated the response of the cell. However, information content was not simply related to firing rate, as there was usually significant information content in the neuronal responses to phase differences that elicited the minimum firing rate. In general, burst responses (impulse intervals <4 ms) did not convey more information than that conveyed by the total response. The contribution to the cumulative information of the response in successive 100-ms segments decreased over the course of the 1-s stimulus. The ratio of information transmitted at 200 ms to that transmitted over the full second had a median of 0.30 while the ratio of 500 ms to 1 s was 0.68.

L1 ANSWER 3 OF 128 MEDLINE
AN 2002200218 MEDLINE
DN 21930716 PubMed ID: 11933064
TI Hamming distance geometry of a protein conformational space: application to the clustering of a 4-ns molecular dynamics trajectory of the HIV-1 integrase catalytic core.
AU Laboulais Cyril; Ouali Mohammed; Le Bret Marc; Gabarro-Arpa Jacques
CS LBPA, CNRS UMR 8532, Ecole Normale Superieure de Cachan, Cachan, France.
SO PROTEINS, (2002 May 1) 47 (2) 169-79.
Journal code: 8700181. ISSN: 1097-0134.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200204
ED Entered STN: 20020405
Last Updated on STN: 20020424
Entered Medline: 20020423
AB Protein structures can be encoded into binary **sequences** (Gabarro-Arpa et al., Comput Chem 2000;24:693-698) these are used to define a Hamming distance in conformational space: the distance between two different molecular conformations is the number of different **bits** in their **sequences**. Each bit in the **sequence** arises from a partition of conformational space in two halves. Thus, the information encoded in the binary **sequences** is also used to characterize the regions of conformational space visited by the system. We apply this distance and their associated geometric structures to the clustering and analysis of conformations sampled during a 4-ns molecular dynamics simulation of the HIV-1 integrase catalytic core. The cluster analysis of the simulation shows a division of the trajectory into two segments of 2.6 and 1.4 ns length, which are qualitatively different: the data points to the fact that equilibration is only reached at the end of the first segment. The Hamming distance is compared also to the r.m.s. deviation measure. The analysis of the cases studied so far shows that under the same conditions the two measures behave quite differently, and that the Hamming distance appears to be more robust than the r.m.s. deviation.

L1 ANSWER 4 OF 128 MEDLINE
AN 2001685242 MEDLINE
DN 21588415 PubMed ID: 11731537
TI Variability and information in a neural code of the cat lateral geniculate nucleus.
AU Liu R C; Tzonev S; Rebrik S; Miller K D
CS Keck Center for Integrative Neuroscience and Department of Physiology, University of California, San Francisco, California 94143-0444, USA.. liu@phy.ucsf.edu
NC R01-EY-13595 (NEI)
R01-NS-33787 (NINDS)
SO JOURNAL OF NEUROPHYSIOLOGY, (2001 Dec) 86 (6) 2789-806.

Journal code: 0375404. ISSN: 0022-3077.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200201
ED Entered STN: 20011204
Last Updated on STN: 20020129
Entered Medline: 20020128
AB A central theme in neural coding concerns the role of response variability and noise in determining the information transmission of neurons. This issue was investigated in single cells of the lateral geniculate nucleus of barbiturate-anesthetized cats by quantifying the degree of precision in and the information transmission properties of individual spike train responses to full field, binary (bright or dark), flashing stimuli. We found that neuronal responses could be highly reproducible in their spike timing (approximately 1-2 ms standard deviation) and spike count (approximately 0.3 ratio of variance/mean, compared with 1.0 expected for a Poisson process). This degree of precision only became apparent when an adequate length of the stimulus **sequence** was specified to determine the neural response, emphasizing that the variables relevant to a cell's response must be controlled to observe the cell's intrinsic response precision. Responses could carry as much as 3.5 **bits** /spike of information about the stimulus, a rate that was within a factor of two of the limit the spike train could transmit. Moreover, there appeared to be little sign of redundancy in coding: on average, longer response **sequences** carried at least as much information about the stimulus as would be obtained by adding together the information carried by shorter response **sequences** considered independently. There also was no direct evidence found for synergy between response **sequences**. These results could largely, but not entirely, be explained by a simple model of the response in which one filters the stimulus by the cell's impulse response kernel, thresholds the result at a fairly high level, and incorporates a postspike refractory period.

L1 ANSWER 5 OF 128 MEDLINE
AN 2001680718 MEDLINE
DN 21583888 PubMed ID: 11726698
TI Strong minor groove base conservation in **sequence** logos implies DNA distortion or base flipping during replication and transcription initiation.
AU Schneider T D
CS National Cancer Institute at Frederick, Laboratory of Experimental and Computational Biology, Building 469, PO Box B, Frederick, MD 21702-1201, USA.. toms@ncifcrf.gov
SO NUCLEIC ACIDS RESEARCH, (2001 Dec 1) 29 (23) 4881-91.
Journal code: 0411011. ISSN: 1362-4962.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200112
ED Entered STN: 20011203
Last Updated on STN: 20020123
Entered Medline: 20011211
AB The **sequence** logo for DNA binding sites of the bacteriophage P1 replication protein RepA shows unusually high **sequence** conservation (approximately 2 **bits**) at a minor groove that faces RepA. However, B-form DNA can support only 1 bit of **sequence** conservation via contacts into the minor groove. The high conservation in RepA sites therefore implies a distorted DNA helix with direct or indirect contacts to the protein. Here I show that a high minor groove conservation signature also appears in **sequence** logos of sites for other replication origin binding proteins (Rts1, DnaA, P4 alpha, EBNA1, ORC) and promoter binding proteins (sigma(70), sigma(D) factors). This finding

implies that DNA binding proteins generally use non-B-form DNA distortion such as base flipping to initiate replication and transcription.

=> d his

(FILE 'HOME' ENTERED AT 17:01:59 ON 04 SEP 2002)

FILE 'MEDLINE, BIOSIS' ENTERED AT 17:02:09 ON 04 SEP 2002
L1 128 S SEQUENCE AND BITS

=> s l1 and (DNA or protein or polypeptide or polynucleotide or amino acid or nucleotide)
L2 67 L1 AND (DNA OR PROTEIN OR POLYPEPTIDE OR POLYNUCLEOTIDE OR AMINO ACID OR NUCLEOTIDE)

=> duplicate remove l2
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE? Y/(N):n
PROCESSING COMPLETED FOR L2
L3 42 DUPLICATE REMOVE L2 (25 DUPLICATES REMOVED)

=> d 1-10 bib ab

L3 ANSWER 1 OF 42 MEDLINE DUPLICATE 1
AN 2002200218 MEDLINE
DN 21930716 PubMed ID: 11933064
TI Hamming distance geometry of a **protein** conformational space: application to the clustering of a 4-ns molecular dynamics trajectory of the HIV-1 integrase catalytic core.
AU Laboulais Cyril; Ouali Mohammed; Le Bret Marc; Gabarro-Arpa Jacques
CS LBPA, CNRS UMR 8532, Ecole Normale Superieure de Cachan, Cachan, France.
SO PROTEINS, (2002 May 1) 47 (2) 169-79.
Journal code: 8700181. ISSN: 1097-0134.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200204
ED Entered STN: 20020405
Last Updated on STN: 20020424
Entered Medline: 20020423
AB **Protein** structures can be encoded into binary **sequences** (Gabarro-Arpa et al., Comput Chem 2000;24:693-698) these are used to define a Hamming distance in conformational space: the distance between two different molecular conformations is the number of different **bits** in their **sequences**. Each bit in the **sequence** arises from a partition of conformational space in two halves. Thus, the information encoded in the binary **sequences** is also used to characterize the regions of conformational space visited by the system. We apply this distance and their associated geometric structures to the clustering and analysis of conformations sampled during a 4-ns molecular dynamics simulation of the HIV-1 integrase catalytic core. The cluster analysis of the simulation shows a division of the trajectory into two segments of 2.6 and 1.4 ns length, which are qualitatively different: the data points to the fact that equilibration is only reached at the end of the first segment. The Hamming distance is compared also to the r.m.s. deviation measure. The analysis of the cases studied so far shows that under the same conditions the two measures behave quite differently, and that the Hamming distance appears to be more robust than the r.m.s. deviation.

L3 ANSWER 2 OF 42 MEDLINE
AN 2001271179 MEDLINE

DN 21228920 PubMed ID: 11330165
TI Bytes and **bits** meet biotech.
AU Sherrid P
SO US NEWS AND WORLD REPORT, (2001 Apr 16) 130 (15) 32-4.
Journal code: 9877797. ISSN: 0041-5537.
CY United States
DT News Announcement
LA English
FS Health
EM 200105
ED Entered STN: 20010529
Last Updated on STN: 20010529
Entered Medline: 20010521

L3 ANSWER 3 OF 42 MEDLINE DUPLICATE 2
AN 2001680718 MEDLINE
DN 21583888 PubMed ID: 11726698
TI Strong minor groove base conservation in **sequence** logos implies **DNA** distortion or base flipping during replication and transcription initiation.
AU Schneider T D
CS National Cancer Institute at Frederick, Laboratory of Experimental and Computational Biology, Building 469, PO Box B, Frederick, MD 21702-1201, USA.. toms@ncifcrf.gov
SO NUCLEIC ACIDS RESEARCH, (2001 Dec 1) 29 (23) 4881-91.
Journal code: 0411011. ISSN: 1362-4962.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200112
ED Entered STN: 20011203
Last Updated on STN: 20020123
Entered Medline: 20011211
AB The **sequence** logo for **DNA** binding sites of the bacteriophage P1 replication **protein** RepA shows unusually high **sequence** conservation (approximately 2 **bits**) at a minor groove that faces RepA. However, B-form **DNA** can support only 1 bit of **sequence** conservation via contacts into the minor groove. The high conservation in RepA sites therefore implies a distorted **DNA** helix with direct or indirect contacts to the **protein** . Here I show that a high minor groove conservation signature also appears in **sequence** logos of sites for other replication origin binding **proteins** (Rts1, DnaA, P4 alpha, EBNA1, ORC) and promoter binding **proteins** (sigma(70), sigma(D) factors). This finding implies that **DNA** binding **proteins** generally use non-B-form **DNA** distortion such as base flipping to initiate replication and transcription.

L3 ANSWER 4 OF 42 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 2002:212005 BIOSIS
DN PREV200200212005
TI Characterization and phylogenetic analyses of novel psychrophiles isolated from an Antarctic lake and a Greenland ice core.
AU Sheridan, P. P. (1); Loveland-Curtze, J. (1); Miteva, V. (1); Brenchley, J. E. (1)
CS (1) Pennsylvania State University, University Park, PA USA
SO Abstracts of the General Meeting of the American Society for Microbiology, (2001) Vol. 101, pp. 435. http://www.asmusa.org/mtgsrc/generalmeeting.htm. print.
Meeting Info.: 101st General Meeting of the American Society for Microbiology Orlando, FL, USA May 20-24, 2001
ISSN: 1060-2011.
DT Conference
LA English

AB Our research interest is focused on isolating novel psychrophiles and characterizing their cold-active enzymes. Gram positive psychrophiles have been obtained from such diverse environments as the Dry Valleys of Antarctica, deep ocean sediments, Siberian permafrost, and whey-enriched Pennsylvania farmland, indicating that these organisms are pan-globally distributed. Recently, we have isolated and characterized two novel psychrophilic Gram positive bacteria from geographically distant environments. The first organism, designated LV3, was isolated from a lake located South of the Miers and Adams glaciers, near the McMurdo Ice Shelf, Antarctica. This red-pigmented organism grew well at 0degreeC, 5degreeC, 10degreeC, and 18degreeC in a wide variety of media but not at higher temperatures. LV3 contains ornithine as the diamino acid in its cell wall linkage. Phylogenetic analysis of isolate LV3's 16S rRNA gene **sequence** indicated that it was related to, but distinct from, organisms belonging to the genera Leifsonia, Corynebacterium, and Curtobacterium, and may represent a new genus. The second organism, designated GIC6, was isolated from a Greenland ice core utilizing a novel sampling method incorporating sterile keyhole drill **bits**. The yellow-pigmented isolate grew well at 18degreeC and 25degreeC and slowly at 10degreeC, but no growth was seen at 37degreeC. Isolate GIC6 was most closely related to strains 301, 312, 801, and NS/4 of the genus Frigoribacterium, based on phylogenetic analysis of its 16S rRNA gene **sequence**. Not only are psychrophilic strains found in many different genera of the high G+C Gram positive bacteria, but these organisms are also distributed globally. This may reflect either reduced rate of change in the 16S rRNA genes of these organisms, or an efficient global bacterial dispersal mechanism.

L3 ANSWER 5 OF 42 MEDLINE
AN 2001555878 MEDLINE
DN 21488553 PubMed ID: 11601857
TI Anatomy of Escherichia coli ribosome binding sites.
AU Shultzaberger R K; Bucheimer R E; Rudd K E; Schneider T D
CS University of Maryland, College Park, 20742, USA.
NC GM58560 (NIGMS)
SO JOURNAL OF MOLECULAR BIOLOGY, (2001 Oct 12) 313 (1) 215-28.
Journal code: 2985088R. ISSN: 0022-2836.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200112
ED Entered STN: 20011017
Last Updated on STN: 20020122
Entered Medline: 20011204
AB During translational initiation in prokaryotes, the 3' end of the 16S rRNA binds to a region just upstream of the initiation codon. The relationship between this Shine-Dalgarno (SD) region and the binding of ribosomes to translation start-points has been well studied, but a unified mathematical connection between the SD, the initiation codon and the spacing between them has been lacking. Using information theory, we constructed a model that treats these three components uniformly by assigning to the SD and the initiation region (IR) conservations in **bits** of information, and by assigning to the spacing an uncertainty, also in **bits**. To build the model, we first aligned the SD region by maximizing the information content there. The ease of this process confirmed the existence of the SD pattern within a set of 4122 reviewed and revised Escherichia coli gene starts. This large data set allowed us to show graphically, by **sequence** logos, that the spacing between the SD and the initiation region affects both the SD site conservation and its pattern. We used the aligned SD, the spacing, and the initiation region to model ribosome binding and to identify gene starts that do not conform to the ribosome binding site model. A total of 569 experimentally proven starts are more conserved (have higher information content) than the full set of revised starts, which probably reflects an experimental bias

against the detection of gene products that have inefficient ribosome binding sites. Models were refined cyclically by removing non-conforming weak sites. After this procedure, models derived from either the original or the revised gene start annotation were similar. Therefore, this information theory-based technique provides a method for easily constructing biologically sensible ribosome binding site models. Such models should be useful for refining gene-start predictions of any sequenced bacterial genome.

L3 ANSWER 6 OF 42 MEDLINE DUPLICATE 3
AN 2001080854 MEDLINE
DN 20553462 PubMed ID: 11099383
TI Searching the **protein** structure databank with weak **sequence** patterns and structural constraints.
AU Jonassen I; Eidhammer I; Grindhaug S H; Taylor W R
CS Department of Informatics, University of Bergen, Hoyteknologisenteret (P.B. 7800), Bergen, N-5020, Norway.
SO JOURNAL OF MOLECULAR BIOLOGY, (2000 Dec 8) 304 (4) 599-619.
Journal code: 2985088R. ISSN: 0022-2836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200101
ED Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20010111
AB A method is described in which **proteins** that match PROSITE patterns are filtered by the root-mean-square deviation of the local 3D structures of the probe and target over the pattern components. This was found to increase the discrimination between true and false members of the **protein** family but was dependent on how unique the structural features in the pattern were compared to equivalent fragments extracted from the structure databank (for example; if the pattern fell in an alpha-helix, then discrimination was poor.) We then generalised the **sequence** patterns (by widening the range of **amino acid** residues allowed at each position) and monitored how well the structural information helped retain specificity. While the discrimination of the pure **sequence** pattern had generally disappeared at information content values less than ten **bits**, the discrimination of the combined **sequence** structure probe remained high at this point before following a similar decay. The displacement between these curves indicates that the structural component is, on average, equivalent to about ten **bits**. The **sequence** patterns were also filtered using the structure comparison program SAP, giving a global, rather than local "view" of the **proteins**. This allowed the information content of the **sequence** patterns to become even less specific but raised problems of whether some **proteins** encountered with the same fold but no PROSITE pattern should constitute family members.

L3 ANSWER 7 OF 42 MEDLINE DUPLICATE 4
AN 2000247403 MEDLINE
DN 20247403 PubMed ID: 10783485
TI Cholinergic regulation of biological hydrodynamics.
AU Axelsson S
CS Department of Obstetrics and Gynaecology, Swedish University of Agricultural Sciences, Uppsala, Sweden.
SO MEDICAL HYPOTHESES, (2000 Mar) 54 (3) 444-7.
Journal code: 7505668. ISSN: 0306-9877.
CY SCOTLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English

FS Priority Journals
EM 200006
ED Entered STN: 20000616
Last Updated on STN: 20000616
Entered Medline: 20000608
AB The structures of biological life are formed in water. Their function depends on changes in the entropy of water. It is regulated by the cholinergic system. The initiating event is the ChE-splitting of water with liberation of free protons. They will draw electrons from the fairly inert dioxygen. The induced oxygen reactivity will give liberation and transfer of electrons and hydrodynamic pH-dependent changes in **protein** configurations. A multitude of sub-systems will be activated. The **sequence** of events normally ends with the formation of water, thus preventing uncontrolled radical chain reactions. Cholinergic receptors appear as restricting units of the general disordering entropy tendency. ChE-induced hydrodynamics is propagated to the inner of cells by the water soluble protons and the electrolytes. Especially Ca appear to have a strong influence on the hydrodynamic dipole moment of water. Because water is an integral structure of **DNA** genetics also will be influenced. Conditions caused by deprivation of oxygen or of reactive oxygen and disorders by hyperactivity and inactivity are briefly discussed. The CNS takes the shape of a large-scale quantum computer with a function far beyond our ability of immediate perception. The atomic nuclear proportions of quantum **bits** (qubits) will admit the functional one-cell unit of immune memory cells. Cholinergic hydrodynamics appear to substantiate the much discussed chaos theory.

L3 ANSWER 8 OF 42 BIOSIS
AN 2001:292598 BIOSIS
DN PREV200100292598
TI A new approach for the detection of control sites in **DNA sequences**.
AU Mahon, Garry A. (1); Dicato, Mario A.
CS (1) Research on Cancer and Blood Disorders, Centre Hospitalier, Luxembourg, Luxembourg Luxembourg
SO Blood, (November 16, 2000) Vol. 96, No. 11 Part 2, pp. 181b. print.
Meeting Info.: 42nd Annual Meeting of the American Society of Hematology San Francisco, California, USA December 01-05, 2000 American Society of Hematology
. ISSN: 0006-4971.
DT Conference
LA English
SL English
AB With the rapidly rising volume of reported **DNA sequences**, there is lively interest in automatic methods for detecting control sites or other short, specific sub-**sequences**. We have developed an approach by which the statistical analysis of a reference set of **sequences** of a particular type of site allows one or more equations to be defined. If such an equation is satisfied by a new **sequence** then it is highly likely that the **sequence** corresponds to a site of the particular type. The definition of the equations makes use of the properties of the eigenvalues and eigenvectors of the covariance matrix of the suitably encoded **sequences**. In particular, the existence of one or more zero eigenvalues implies the existence of one or more such equations. The approach is illustrated with the **sequences** of 173 promoters recognised by human RNA polymerase II. Sub-**sequences** of 25 bases around the TATA site were extracted. Two **bits** were used to encode each base and the covariance matrix of the resulting 50 variables was calculated. The eigenvalues of this matrix ranged from 0.787 down to 0.035. The eigenvalue of 0.035 (almost zero) means that there is an equation which is (almost) satisfied by all the promoters in the set. A new **sequence** which (almost) satisfies this equation may be regarded as a putative promoter. This approach can be used for scanning any **DNA** database looking

for **sequences** of particular interest.

L3 ANSWER 9 OF 42 MEDLINE DUPLICATE 5
AN 2000234588 MEDLINE
DN 20234588 PubMed ID: 10774555
TI Free energy and information contents of conformons in **proteins** and **DNA**.
AU Ji S
CS Department of Pharmacology and Toxicology, Rutgers University, Piscataway, NJ 08855, USA.. sji@eohsi.rutgers.edu
SO BIOSYSTEMS, (2000 Jan) 54 (3) 107-30. Ref: 145
Journal code: 0430773. ISSN: 0303-2647.
CY Ireland
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 200006
ED Entered STN: 20000622
Last Updated on STN: 20000622
Entered Medline: 20000613
AB **Sequence**-specific conformational strains (SSCS) of biopolymers that carry free energy and genetic information have been called conformons, a term coined independently by two groups over two and a half decades ago [Green, D.E., Ji, S., 1972. The electromechanochemical model of mitochondrial structure and function. In: Schultz, J., Cameron, B.F. (Eds.), Molecular Basis of Electron Transport. Academic Press, New York, pp. 1-44; Volkenstein, M.V., 1972. The Conformon. J. Theor. Biol. 34, 193-195]. Conformons provide the molecular mechanisms necessary and sufficient to account for all biological processes in the living cell on the molecular level in principle--including the origin of life, enzymic catalysis, control of gene expression, oxidative phosphorylation, active transport, and muscle contraction. A clear example of SSCS is provided by SIDD (strain-induced duplex destabilization) in **DNA** recently reported by Benham [Benham, C.J., 1996a. Duplex destabilization in superhelical **DNA** is predicted to occur at specific transcriptional regulatory regions. J. Mol. Biol. 255, 425-434; Benham, C.J., 1996b. Computation of **DNA** structural variability--a new predictor of **DNA** regulatory regions. CABIOS 12(5), 375-381]. Experimental as well as theoretical evidence indicates that conformons in **proteins** carry 8-16 kcal/mol of free energy and 40-200 **bits** of information, while those in **DNA** contain 500-2500 kcal/mol of free energy and 200-600 **bits** of information. The similarities and differences between conformons and solitons have been analyzed on the basis of the generalized Franck-Condon principle [Ji, S., 1974a. A general theory of ATP synthesis and utilization. Ann. N.Y. Acad. Sci. 227, 211-226; Ji, S., 1974b. Energy and negentropy in enzymic catalysis. Ann. N.Y. Acad. Sci. 227, 419-437]. To illustrate a practical application, the conformon theory was applied to the molecular-clamp model of **DNA** gyrase proposed by Berger and Wang [Berger, J.M., Wang, J.C., 1996. Recent developments in **DNA** topoisomerases II structure and mechanism. Curr. Opin. Struct. Biol. 6(1), 84-90], leading to the proposal of an eight-step molecular mechanism for the action of the enzyme. Finally, a set of experimentally testable predictions has been formulated on the basis of the conformon theory.

L3 ANSWER 10 OF 42 MEDLINE
AN 2001646608 MEDLINE
DN 21557032 PubMed ID: 11700586
TI Biological **sequence** compression algorithms.
AU Matsumoto T; Sadakane K; Imai H
CS Department of Information Science, University of Tokyo,7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.. toshikom@is.s.u-tokyo.ac.jp
SO GENOME INFORMATICS SERIES, (2000) 11 43-52.

Journal code: 9717234. ISSN: 0919-9454.
CY Japan
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200112
ED Entered STN: 20011112
Last Updated on STN: 20020124
Entered Medline: 20011231
AB Today, more and more **DNA sequences** are becoming available. The information about **DNA sequences** are stored in molecular biology databases. The size and importance of these databases will be bigger and bigger in the future, therefore this information must be stored or communicated efficiently. Furthermore, **sequence** compression can be used to define similarities between biological **sequences**. The standard compression algorithms such as gzip or compress cannot compress **DNA sequences**, but only expand them in size. On the other hand, CTW (Context Tree Weighting Method) can compress **DNA sequences** less than two **bits** per symbol. These algorithms do not use special structures of biological **sequences**. Two characteristic structures of **DNA sequences** are known. One is called palindromes or reverse complements and the other structure is approximate repeats. Several specific algorithms for **DNA sequences** that use these structures can compress them less than two **bits** per symbol. In this paper, we improve the CTW so that characteristic structures of **DNA sequences** are available. Before encoding the next symbol, the algorithm searches an approximate repeat and palindrome using hash and dynamic programming. If there is a palindrome or an approximate repeat with enough length then our algorithm represents it with length and distance. By using this preprocessing, a new program achieves a little higher compression ratio than that of existing **DNA**-oriented compression algorithms. We also describe new compression algorithm for **protein sequences**.

=> d 11-20 bib ab

L3 ANSWER 11 OF 42 MEDLINE DUPLICATE 6
AN 1999238185 MEDLINE
DN 99238185 PubMed ID: 10223669
TI Significantly lower entropy estimates for natural **DNA sequences**.
AU Loewenstern D; Yianilos P N
CS NEC Research Institute, Princeton, New Jersey 08540, USA..
davel@research.nj.nec.com
SO JOURNAL OF COMPUTATIONAL BIOLOGY, (1999 Spring) 6 (1) 125-42.
Journal code: 9433358. ISSN: 1066-5277.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199906
ED Entered STN: 19990618
Last Updated on STN: 19990618
Entered Medline: 19990607
AB If **DNA** were a random string over its alphabet A, C, G, T , an optimal code would assign two **bits** to each **nucleotide**. **DNA** may be imagined to be a highly ordered, purposeful molecule, and one might therefore reasonably expect statistical models of its string representation to produce much lower entropy estimates. Surprisingly, this has not been the case for many natural **DNA sequences**, including portions of the human genome. We introduce a new statistical model (compression algorithm), the strongest reported to date, for naturally occurring **DNA sequences**. Conventional

techniques code a **nucleotide** using only slightly fewer **bits** (1.90) than one obtains by relying only on the frequency statistics of individual **nucleotides** (1.95). Our method in some cases increases this gap by more than fivefold (1.66) and may lead to better performance in microbiological pattern recognition applications. One of our main contributions, and the principle source of these improvements, is the formal inclusion of inexact match information in the model. The existence of matches at various distances forms a panel of experts which are then combined into a single prediction. The structure of this combination is novel and its parameters are learned using Expectation Maximization (EM). Experiments are reported using a wide variety of **DNA sequences** and compared whenever possible with earlier work. Four reasonable notions for the string distance function used to identify near matches, are implemented and experimentally compared. We also report lower entropy estimates for coding regions extracted from a large collection of nonredundant human genes. The conventional estimate is 1.92 **bits**. Our model produces only slightly better results (1.91 **bits**) when considering **nucleotides**, but achieves 1.84-1.87 **bits** when the prediction problem is divided into two stages: (i) predict the next **amino acid**-based on inexact **polypeptide** matches, and (ii) predict the particular codon. Our results suggest that matches at the **amino acid** level play some role, but a small one, in determining the statistical structure of nonredundant coding **sequences**.

L3 ANSWER 12 OF 42 MEDLINE
AN 2000099982 MEDLINE
DN 20099982 PubMed ID: 10636027
TI Progress toward demonstration of a surface based **DNA** computation: a one word approach to solve a model satisfiability problem.
AU Liu Q; Frutos A G; Wang L; Thiel A J; Gillmor S D; Strother C T; Condon A E; Corn R M; Lagally M G; Smith L M
CS Department of Chemistry, University of Wisconsin-Madison, 53706, USA.
SO BIOSYSTEMS, (1999 Oct) 52 (1-3) 25-33.
Journal code: 0430773. ISSN: 0303-2647.
CY Ireland
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200002
ED Entered STN: 20000314
Last Updated on STN: 20000314
Entered Medline: 20000225
AB A multi-base encoding strategy is used in a one word approach to surface-based **DNA** computation. In this designed **DNA** model system, a set of 16 oligonucleotides, each a 16mer, is used with the format 5'-FFFFvvvvvvvvFFFF-3' in which 4-8 **bits** of data are stored in eight central variable ('v') base locations, and the remaining fixed ('F') base locations are used as a word label. The detailed implementations are reported here. In order to achieve perfect discrimination between each oligonucleotide, the efficiency and specificity of hybridization discrimination of the set of 16 oligonucleotides were examined by carrying out the hybridization of each individual fluorescently tagged complement to an array of 16 addressed immobilized oligonucleotides. A series of preliminary hybridization experiments are presented and further studies about hybridization, enzymatic destruction, read out and demonstrations of a SAT problem are forthcoming.

L3 ANSWER 13 OF 42 MEDLINE
AN 1998414046 MEDLINE
DN 98414046 PubMed ID: 9743114
TI From **bits** to bases: computing with **DNA**.
AU Dove A

SO NATURE BIOTECHNOLOGY, (1998 Sep) 16 (9) 830-2.
Journal code: 9604648. ISSN: 1087-0156.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199811
ED Entered STN: 19990106
Last Updated on STN: 19990106
Entered Medline: 19981123

L3 ANSWER 14 OF 42 MEDLINE
AN 1998391762 MEDLINE
DN 98391762 PubMed ID: 9724323
TI Reading **bits** of genetic information: methods for single-**nucleotide** polymorphism analysis.
AU Landegren U; Nilsson M; Kwok P Y
CS Department of Genetics and Pathology, Uppsala University, Se-751 23 Uppsala, Sweden.. ulf.landegren@medgen.uu.se
SO GENOME RESEARCH, (1998 Aug) 8 (8) 769-76. Ref: 53
Journal code: 9518021. ISSN: 1088-9051.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 199809
ED Entered STN: 19981006
Last Updated on STN: 19981006
Entered Medline: 19980923

L3 ANSWER 15 OF 42 MEDLINE DUPLICATE 7
AN 1998375691 MEDLINE
DN 98375691 PubMed ID: 9711873
TI Information analysis of human splice site mutations.
CM Erratum in: Hum Mutat 1999;13(1):82
AU Rogan P K; Faux B M; Schneider T D
CS Department of Human Genetics, Allegheny University of the Health Sciences, Pittsburgh, PA 15212, USA.. progan@pgh.allegheny.edu
NC CA74683 (NCI)
SO HUMAN MUTATION, (1998) 12 (3) 153-71.
Journal code: 9215429. ISSN: 1059-7794.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199810
ED Entered STN: 19981021
Last Updated on STN: 20000303
Entered Medline: 19981015
AB Splice site **nucleotide** substitutions can be analyzed by comparing the individual information contents (Ri, **bits**) of the normal and variant splice junction **sequences** [Rogan and Schneider, 1995]. In the present study, we related splicing abnormalities to changes in Ri values of 111 previously reported splice site substitutions in 41 different genes. Mutant donor and acceptor sites have significantly less information than their normal counterparts. With one possible exception, primary mutant sites with <2.4 **bits** were not spliced. Sites with Ri values > or = 2.4 **bits** but less than the corresponding natural site usually decreased, but did not abolish splicing. Substitutions that produced small changes in Ri probably do not impair splicing and are often polymorphisms. The Ri values of activated cryptic sites were generally comparable to or greater than those of the corresponding natural splice sites. Information analysis revealed

preexisting cryptic splice junctions that are used instead of the mutated natural site. Other cryptic sites were created or strengthened by **sequence** changes that simultaneously altered the natural site. Comparison between normal and mutant splice site Ri values distinguishes substitutions that impair splicing from those which do not, distinguishes null alleles from those that are partially functional, and detects activated cryptic splice sites.

L3 ANSWER 16 OF 42 MEDLINE
AN 1999030754 MEDLINE
DN 99030754 PubMed ID: 9788845
TI Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene.
AU Kannabiran C; Rogan P K; Olmos L; Basti S; Rao G N; Kaiser-Kupfer M; Hejtmancik J F
CS National Eye Institute, NIH, Bethesda, MD 20892-1860, USA.
SO MOLECULAR VISION, (1998 Oct 23) 4 21.
Journal code: 9605351. ISSN: 1090-0535.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199811
ED Entered STN: 19990106
Last Updated on STN: 19990106
Entered Medline: 19981125
AB PURPOSE: Congenital cataracts constitute a morphologically and genetically heterogeneous group of diseases that are a major cause of childhood blindness. Autosomal Dominant Zonular Cataracts with Sutural Opacities (CCZS) have been mapped to chromosome 17q11-q12 near the betaA3A1-crystallin gene (CRYBA1). The betaA3A1-crystallin gene was investigated as the causative gene for the cataracts. METHODS: The betaA3/A1-crystallin gene was sequenced in affected and control individuals. Base changes were confirmed and assayed in additional family members and controls using NlaIII restriction digestion of PCR amplified **DNA sequences**. Base changes were assessed for their effects on splicing by information analysis. RESULTS: The cataracts are associated with a **sequence** change in the 5' (donor) splice site of intron 3: GC(g->a)tgagt. The **sequence** change also creates a new NlaIII site. This base change cosegregates with the cataracts in this family, being present in every affected individual. Conversely, this base change was not seen in 140 chromosomes examined in 70 unaffected and unrelated individuals. Information theory mutational analysis shows that the base change lowers the information content of the splice site from 6.0 to -6.8 **bits**, so that splicing would not be expected to occur at the altered site. CONCLUSIONS: Taken together, these observations suggest that the observed mutation might be causally related to the cataracts in this family.

L3 ANSWER 17 OF 42 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1998:518293 BIOSIS
DN PREV199800518293
TI Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene.
AU Kannabiran, Chitra; Rogan, Peter K.; Olmos, Lisa; Basti, Surrendra; Rao, Gullapalli N.; Kaiser-Kupfer, Muriel; Hejtmancik, J. Fielding (1)
CS (1) OGCSB/NEI/NIH, Building 10, Room 10B10, 10 Center Dr Msc 1860, Bethesda, MD 20892-1860 USA
SO Molecular Vision, (Oct. 23, 1998) Vol. 4, No. 21 CITED OCT. 28, 1998, pp. NO PAGINATION. http://www.emory.edu/molvis/v4/p21.
ISSN: 1090-0535.
DT Article
LA English
AB Purpose: Congenital cataracts constitute a morphologically and genetically heterogeneous group of diseases that are a major cause of childhood

blindness. Autosomal Dominant Zonular Cataracts with Sutural Opacities (CCZS) have been mapped to chromosome 17q11-q12 near the betaA3A1-crystallin gene (CRYBA1). The betaA3A1-crystallin gene was investigated as the causative gene for the cataracts. Methods: The betaA3/A1-crystallin gene was sequenced in affected and control individuals. Base changes were confirmed and assayed in additional family members and controls using NlaIII restriction digestion of PCR amplified **DNA sequences**. Base changes were assessed for their effects on splicing by information analysis. Results: The cataracts are associated with a **sequence** change in the 5' (donor) splice site of intron 3: GC(g->a)tgagt. The **sequence** change also creates a new NlaIII site. This base change cosegregates with the cataracts in this family, being present in every affected individual. Conversely, this base change was not seen in 140 chromosomes examined in 70 unaffected and unrelated individuals. Information theory mutational analysis shows that the base change lowers the information content of the splice site from 6.0 to -6.8 **bits**,so that splicing would not be expected to occur at the altered site. Conclusions: Taken together, these observations suggest that the observed mutation might be causally related to the cataracts in this family.

L3 ANSWER 18 OF 42 MEDLINE DUPLICATE 8
AN 1998097036 MEDLINE
DN 98097036 PubMed ID: 9441280
TI Demonstration of a word design strategy for **DNA** computing on surfaces.
AU Frutos A G; Liu Q; Thiel A J; Sanner A M; Condon A E; Smith L M; Corn R M
CS Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA.. corn@chem.wosc.edu
SO NUCLEIC ACIDS RESEARCH, (1997 Dec 1) 25 (23) 4748-57.
Journal code: 0411011. ISSN: 0305-1048.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199801
ED Entered STN: 19980130
Last Updated on STN: 19980130
Entered Medline: 19980120
AB A strategy for **DNA** computing on surfaces using linked sets of ' **DNA** words' that are short oligonucleotides (16mers) is proposed. The 16mer words have the format 5'-FFFFvvvvvvvvFFFF-3' in which 4-8 **bits** of data are stored in 8 variable ('v') base locations, and the remaining fixed ('F') base locations are used as a word label. Using a template and map strategy, a set of 108 8mers each of which possesses at least a 4 base mismatch with the complements to all the other members of the set (4bm complements) are identified for use as a variable base **sequence** set. In addition, sets of 4 and 12 word labels of the form ABCD....DCBA that are respectively 8bm and 6bm complements with each other are identified. The 16mers are chosen to have a G/C content of 50% in order to make the thermodynamic stability of the perfectly matched hybridized **DNA** duplexes similar; a simple pairwise additive method is used to estimate the perfect match and mismatch hybridization thermodynamics. A series of preliminary experiments are presented that use small arrays of 16mers attached to chemically modified gold surfaces and fluorescently labeled complements to study the hybridization adsorption and enzymatic manipulation of the oligonucleotides.

L3 ANSWER 19 OF 42 MEDLINE DUPLICATE 9
AN 1998114562 MEDLINE
DN 98114562 PubMed ID: 9446751
TI Information content of individual genetic **sequences**.
AU Schneider T D
CS National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Mathematical Biology, P.O. Box B, Frederick, MD

21702-1201, USA.. toms@ncifcrf.gov
SO JOURNAL OF THEORETICAL BIOLOGY, (1997 Dec 21) 189 (4) 427-41.
Journal code: 0376342. ISSN: 0022-5193.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199803
ED Entered STN: 19980410
Last Updated on STN: 20000413
Entered Medline: 19980330
AB Related genetic **sequences** having a common function can be described by Shannon's information measure and depicted graphically by a **sequence** logo. Though useful for many purposes, **sequence** logos only show the average **sequence** conservation, and inferring the conservation for individual **sequences** is difficult. This limitation is overcome by the individual information (R i) technique described here. The method begins by generating a weight matrix from the frequencies of each **nucleotide** or **amino acid** at each position of the aligned **sequences**. This matrix is then applied to the **sequences** themselves to determine the **sequence** conservation of each individual **sequence**. The matrix is unique because the average of these assignments is the total **sequence** conservation, ad there is only one way to construct such a matrix. For binding sites on **polynucleotides**, the weight matrix has a natural cut off that distinguishes functional **sequences** from other **sequences**. R i values are on an absolute scale measured in **bits** of information so the conservation of different biological functions can be compared with one another. The matrix can be used to rank-order the **sequences**, to search for new **sequences**, to compare **sequences** to other quantitative data such as binding energy or distance between binding sites, to distinguish mutations from polymorphisms, to design **sequences** of a given strength, and to detect errors in databases. The R i method has been used to identify previously undescribed but experimentally verified **DNA** binding sites. The individual information distribution was determined for E. coli ribosome binding sites, bacterial Fis binding sites, and human donor and acceptor splice junctions, among others. The distributions demonstrate clearly that the consensus **sequence** is highly unusual, and hence is a poor method to describe naturally occurring binding sites.

L3 ANSWER 20 OF 42 MEDLINE DUPLICATE 10
AN 97081098 MEDLINE
DN 97081098 PubMed ID: 8922380
TI Dictyostelium discoideum cells lacking the 34,000-dalton actin-binding **protein** can grow, locomote, and develop, but exhibit defects in regulation of cell structure and movement: a case of partial redundancy.
AU Rivero F; Furukawa R; Noegel A A; Fechheimer M
CS Max-Planck-Institute for Biochemistry, Martinsried, Germany.
SO JOURNAL OF CELL BIOLOGY, (1996 Nov) 135 (4) 965-80.
Journal code: 0375356. ISSN: 0021-9525.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-U32112; GENBANK-Z50156
EM 199701
ED Entered STN: 19970128
Last Updated on STN: 19990129
Entered Medline: 19970109
AB Cells lacking the Dictyostelium 34,000-D actin-bundling **protein**, a calcium-regulated actin cross-linking **protein**, were created to probe the function of this **polypeptide** in living cells. Gene

replacement vectors were constructed by inserting either the UMP synthase or hygromycin resistance cassette into cloned 4-kb genomic **DNA** containing **sequences** encoding the 34-kD **protein**. After transformation and growth under appropriate selection, cells lacking the **protein** were analyzed by PCR analyses on genomic **DNA**, Northern blotting, and Western blotting. Cells lacking the 34-kD **protein** were obtained in strains derived from AX2 and AX3. Growth, pinocytosis, morphogenesis, and expression of developmentally regulated genes is normal in cells lacking the 34-kD **protein**. In chemotaxis studies, 34-kD- cells were able to locomote and orient normally, but showed an increased persistence of motility. The 34-kD- cells also lost **bits** of cytoplasm during locomotion. The 34-kD- cells exhibited either an excessive number of long and branched filopodia, or a decrease in filopodial length and an increase in the total number of filopodia per cell depending on the strain. Reexpression of the 34-kD **protein** in the AX2-derived strain led to a "rescue" of the defect in the persistence of motility and of the excess numbers of long and branched filopodia, demonstrating that these defects result from the absence of the 34-kD **protein**. We explain the results through a model of partial functional redundancy. Numerous other actin cross-linking **proteins** in Dictyostelium may be able to substitute for some functions of the 34-kD **protein** in the 34-kD cells. The observed phenotype is presumed to result from functions that cannot be adequately supplanted by a substitution of another actin cross-linking **protein**. We conclude that the 34-kD actin-bundling **protein** is not essential for growth, but plays an important role in dynamic control of cell shape and cytoplasmic structure.

```
=> FIL STNGUIDE
COST IN U.S. DOLLARS                                SINCE FILE      TOTAL
                                                       ENTRY    SESSION
FULL ESTIMATED COST                                    21.07      21.28

FILE 'STNGUIDE' ENTERED AT 17:13:38 ON 04 SEP 2002
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY, JAPAN SCIENCE
AND TECHNOLOGY CORPORATION, AND FACHINFORMATIONSZENTRUM KARLSRUHE

FILE CONTAINS CURRENT INFORMATION.
LAST RELOADED: Aug 30, 2002 (20020830/UP).

=> file .pub
COST IN U.S. DOLLARS                                SINCE FILE      TOTAL
                                                       ENTRY    SESSION
FULL ESTIMATED COST                                     0.24      21.52

FILE 'MEDLINE' ENTERED AT 17:15:57 ON 04 SEP 2002

FILE 'BIOSIS' ENTERED AT 17:15:57 ON 04 SEP 2002
COPYRIGHT (C) 2002 BIOLOGICAL ABSTRACTS INC.(R)

=> d 21-30 bib ab

L3   ANSWER 21 OF 42     MEDLINE                          DUPLICATE 11
AN   96397705     MEDLINE
DN   96397705    PubMed ID: 8804598
TI   The Shannon information entropy of protein sequences.
AU   Strait B J; Dewey T G
CS   Department of Chemistry, University of Denver, Colorado 80208, USA.
NC   1R15GM51019 (NIGMS)
SO   BIOPHYSICAL JOURNAL, (1996 Jul) 71 (1) 148-55.
     Journal code: 0370626. ISSN: 0006-3495.
CY   United States
DT   Journal; Article; (JOURNAL ARTICLE)
```

LA English
FS Priority Journals
EM 199612
ED Entered STN: 19970128
Last Updated on STN: 19970128
Entered Medline: 19961218
AB A comprehensive data base is analyzed to determine the Shannon information content of a **protein sequence**. This information entropy is estimated by three methods: a k-tuplet analysis, a generalized Zipf analysis, and a "Chou-Fasman gambler." The k-tuplet analysis is a "letter" analysis, based on conditional **sequence** probabilities. The generalized Zipf analysis demonstrates the statistical linguistic qualities of **protein sequences** and uses the "word" frequency to determine the Shannon entropy. The Zipf analysis and k-tuplet analysis give Shannon entropies of approximately 2.5 **bits/amino acid**. This entropy is much smaller than the value of 4.18 **bits/amino acid** obtained from the nonuniform composition of **amino acids** in **proteins**. The "Chou-Fasman" gambler is an algorithm based on the Chou-Fasman rules for **protein** structure. It uses both **sequence** and secondary structure information to guess at the number of possible **amino acids** that could appropriately substitute into a **sequence**. As in the case for the English language, the gambler algorithm gives significantly lower entropies than the k-tuplet analysis. Using these entropies, the number of most probable **protein sequences** can be calculated. The number of most probable **protein sequences** is much less than the number of possible **sequences** but is still much larger than the number of **sequences** thought to have existed throughout evolution. Implications of these results for mutagenesis experiments are discussed.

L3 ANSWER 22 OF 42 MEDLINE
AN 96019086 MEDLINE
DN 96019086 PubMed ID: 8579892
TI Putting the **bits** and pieces of the RET proto-oncogene puzzle together.
AU Gagel R F
CS Section of Endocrinology, University of Texas M. D. Anderson Cancer Center, Houston 77030, USA.
SO BONE, (1995 Aug) 17 (2 Suppl) 13S-16S. Ref: 30
Journal code: 8504048. ISSN: 8756-3282.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 199603
ED Entered STN: 19960327
Last Updated on STN: 20000303
Entered Medline: 19960320
AB The RET proto-oncogene has been implicated in the causation of papillary thyroid carcinoma, multiple endocrine neoplasia types 2A (MEN 2A) and 2B (MEN 2B), and Hirschsprung's disease. The mutations in these syndromes can be categorized into activating or inactivating mutations. Activating mutations of a cysteine-rich extracellular region cause enhanced dimerization of the RET tyrosine kinase receptor and autophosphorylation, and are causative for MEN 2A and familial medullary thyroid carcinoma (FMTC). An activating mutation of the tyrosine kinase domain causes increased autophosphorylation but does not affect the state of dimerization. A variety of inactivating mutations of the RET proto-oncogene, which result in defective **protein** formation, are causative for Hirschsprung's disease.

L3 ANSWER 23 OF 42 MEDLINE
AN 95009978 MEDLINE
DN 95009978 PubMed ID: 7925317
TI The renin-angiotensin system.
AU Inagami T
CS Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN 37232.
SO ESSAYS IN BIOCHEMISTRY, (1994) 28 147-64. Ref: 20
Journal code: 0043306. ISSN: 0071-1365.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 199411
ED Entered STN: 19941222
Last Updated on STN: 19941222
Entered Medline: 19941110
AB Unravelling of the molecular mechanisms of the action of RAS has been slow. Nature has been rather stingy in revealing **bits** and pieces of information. Each step of development has depended on the innovation of an appropriate methodology. The uniqueness of the RAS lies in: The function and regulation of the highly specific enzyme renin which specifically catalyses the conversion of the prohormone angiotensinogen to Ang I by an extracellular mechanism. The production of the agonist Ang II takes place in two steps. Ang II and its metabolites exert exceedingly diverse pathophysiological effects, presumably through the complex and multifunctional receptors. The exquisite mechanisms involved in the regulation of renin release and receptor regulation are fascinating. The intricate mechanisms that nature has devised for the checks and balances to maintain steady blood flow and electrolyte balance present a great challenge to biochemists in their attempts to clarify the mechanisms involved at both molecular and cellular levels. In relation to the pathophysiology of hypertension, particularly essential hypertension, there is no question that the RAS plays a pivotal role. Although numerous mechanisms could explain its hypertensinogenic effects, no single mechanism can be identified as the major determinant at the present stage of our knowledge. However, there is an important consensus that the effect of Ang II is manifested slowly at even subpressor doses of Ang II through long-term effects involving remodelling of the cardiovascular and renal system.

L3 ANSWER 24 OF 42 MEDLINE DUPLICATE 12
AN 93257451 MEDLINE
DN 93257451 PubMed ID: 7683911
TI **Sequence** analysis of RNA species synthesized by Q beta replicase without template.
AU Biebricher C K; Luce R
CS Max Planck Institute for Biophysical Chemistry, Gottingen, Federal Republic of Germany.
SO BIOCHEMISTRY, (1993 May 11) 32 (18) 4848-54.
Journal code: 0370623. ISSN: 0006-2960.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199306
ED Entered STN: 19930625
Last Updated on STN: 19960129
Entered Medline: 19930611
AB Q beta replicase amplifies certain short-chained RNA templates autocatalytically with high efficiency. In the absence of extraneously added template, synthesis of new RNA species by Q beta replicase is observed under conditions of high enzyme and substrate concentrations and

after long lag times. Even under identical conditions, different RNA species are produced in different experiments. The **sequences** of several independent template-free products have been determined by cloning their cDNAs into plasmids by a novel cloning procedure. Their **nucleotide** chain lengths are small, ranging from 25 to about 50 **nucleotides**. While their primary **sequences** are unrelated except for the invariant 5'-terminal G and 3'-terminal C clusters, their tentative secondary structures show a common principle: both their plus and minus strands have a stem at the 5' terminus, while the 3' terminus is unpaired. Direct accumulation of sufficient quantities of early template-free synthesis products by Q beta replicase is prevented by the inherent irreproducibility of the synthesis process and by the rapid change of the products during amplification by evolution processes, but large amounts of such RNA can be synthesized in vitro by transcription from the cDNA clones. RNA species produced in template-free reactions replicate much more slowly than the optimized RNA species characterized previously. These experimental results illustrate how biological information can be gained in small **bits** by trial and error.

L3 ANSWER 25 OF 42 MEDLINE DUPLICATE 13
AN 94005759 MEDLINE
DN 94005759 PubMed ID: 8402207
TI Discovering simple **DNA sequences** by the algorithmic significance method.
AU Milosavljevic A; Jurka J
CS Linus Pauling Institute of Science and Medicine, Palo Alto, CA 94306.
SO COMPUTER APPLICATIONS IN THE BIOSCIENCES, (1993 Aug) 9 (4) 407-11. Journal code: 8511758. ISSN: 0266-7061.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199311
ED Entered STN: 19940117
Last Updated on STN: 19980206
Entered Medline: 19931119
AB A new method, 'algorithmic significance', is proposed as a tool for discovery of patterns in **DNA sequences**. The main idea is that patterns can be discovered by finding ways to encode the observed data concisely. In this sense, the method can be viewed as a formal version of the Occam's Razor principle. In this paper the method is applied to discover significantly simple **DNA sequences** . We define **DNA sequences** to be simple if they contain repeated occurrences of certain 'words' and thus can be encoded in a small number of **bits**. Such definition includes minisatellites and microsatellites. A standard dynamic programming algorithm for data compression is applied to compute the minimal encoding lengths of **sequences** in linear time. An electronic mail server for identification of simple **sequences** based on the proposed method has been installed at the Internet address pythia/anl.gov.

L3 ANSWER 26 OF 42 MEDLINE DUPLICATE 14
AN 93247069 MEDLINE
DN 93247069 PubMed ID: 8483166
TI A **protein** alignment scoring system sensitive at all evolutionary distances.
AU Altschul S F
CS National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894.
SO JOURNAL OF MOLECULAR EVOLUTION, (1993 Mar) 36 (3) 290-300. Journal code: 0360051. ISSN: 0022-2844.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Space Life Sciences

EM 199305
ED Entered STN: 19930618
Last Updated on STN: 19930618
Entered Medline: 19930528
AB **Protein sequence** alignments generally are constructed with the aid of a "substitution matrix" that specifies a score for aligning each pair of **amino acids**. Assuming a simple random **protein** model, it can be shown that any such matrix, when used for evaluating variable-length local alignments, is implicitly a "log-odds" matrix, with a specific probability distribution for **amino acid** pairs to which it is uniquely tailored. Given a model of **protein** evolution from which such distributions may be derived, a substitution matrix adapted to detecting relationships at any chosen evolutionary distance can be constructed. Because in a database search it generally is not known a priori what evolutionary distances will characterize the similarities found, it is necessary to employ an appropriate range of matrices in order not to overlook potential homologies. This paper formalizes this concept by defining a scoring system that is sensitive at all detectable evolutionary distances. The statistical behavior of this scoring system is analyzed, and it is shown that for a typical **protein** database search, estimating the originally unknown evolutionary distance appropriate to each alignment costs slightly over two **bits** of information, or somewhat less than a factor of five in statistical significance. A much greater cost may be incurred, however, if only a single substitution matrix, corresponding to the wrong evolutionary distance, is employed.

L3 ANSWER 27 OF 42 MEDLINE
AN 96038977 MEDLINE
DN 96038977 PubMed ID: 7584347
TI Discovering **sequence** similarity by the algorithmic significance method.
AU Milosavljevic A
CS Biological and Medical Research Division, Argonne National Laboratory, Illinois 60439-4833, USA.
SO ISMB, (1993) 1 284-91.
Journal code: 9509125.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199512
ED Entered STN: 19960124
Last Updated on STN: 19980206
Entered Medline: 19951205
AB The minimal-length encoding approach is applied to define concept of **sequence** similarity. A **sequence** is defined to be similar to another **sequence** or to a set of keywords if it can be encoded in a small number of **bits** by taking advantage of common subwords. Minimal-length encoding of a **sequence** is computed in linear time, using a data compression algorithm that is based on a dynamic programming strategy and the directed acyclic word graph data structure. No assumptions about common word ("k-tuple") length are made in advance, and common words of any length are considered. The newly proposed algorithmic significance method provides an exact upper bound on the probability that **sequence** similarity has occurred by chance, thus eliminating the need for any arbitrary choice of similarity thresholds. Preliminary experiments indicate that a small number of keywords can positively identify a **DNA sequence**, which is extremely relevant in the context of partial sequencing by hybridization.

L3 ANSWER 28 OF 42 MEDLINE DUPLICATE 15
AN 93393190 MEDLINE
DN 93393190 PubMed ID: 8379664

TI High-tech breakthrough **DNA** scanner for reading **sequence** and detecting gene mutation. A powerful 1 lb, 20 micron resolution, 16-bit personal scanner (PS) that scans 17" x 14" X-ray film in 48 s, with laser, UV and white light sources.
AU Zeineh J A; Zeineh M M; Zeineh R A
CS University of California, San Diego, La Jolla 92037-1321.
SO APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY, (1993 Jun) 41 (3) 219-31.
Journal code: 8208561. ISSN: 0273-2289.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199310
ED Entered STN: 19931105
Last Updated on STN: 19931105
Entered Medline: 19931020
AB The 17" x 14" X-ray film, gels, and blots are widely used in **DNA** research. However, **DNA** laser scanners are costly and unaffordable for the majority of surveyed biotech scientists who need it. The high-tech breakthrough analytical personal scanner (PS) presented in this report is an inexpensive 1 lb hand-held scanner priced at 2-4% of the bulky and costly 30-95 lb conventional laser scanners. This PS scanner is affordable from an operation budget and biotechnologists, who originate most science breakthroughs, can acquire it to enhance their speed, accuracy, and productivity. Compared to conventional laser scanners that are currently available only through hard-to-get capital-equipment budgets, the new PS scanner offers improved spatial resolution of 20 microns, higher speed (scan up to 17" x 14" molecular X-ray film in 48 s), 1-32,768 gray levels (16-**bits**), student routines, versatility, and, most important, affordability. Its programs image the film, read **DNA sequences** automatically, and detect gene mutation. In parallel to the wide laboratory use of PC computers instead of mainframes, this PS scanner might become an integral part of a PC-PS powerful and cost-effective system where the PS performs the digital imaging and the PC acts on the data.

L3 ANSWER 29 OF 42 MEDLINE DUPLICATE 16
AN 93206356 MEDLINE
DN 93206356 PubMed ID: 8456498
TI Regulating genes by packaging domains: **bits** of heterochromatin in euchromatin?.
AU Shaffer C D; Wallrath L L; Elgin S C
CS Department of Biology, Washington University, St Louis, MO 63130.
NC GM14516 (NIGMS)
HD23844 (NICHD)
SO TRENDS IN GENETICS, (1993 Feb) 9 (2) 35-7.
Journal code: 8507085. ISSN: 0168-9525.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199304
ED Entered STN: 19930507
Last Updated on STN: 19990129
Entered Medline: 19930416

L3 ANSWER 30 OF 42 MEDLINE DUPLICATE 17
AN 92276334 MEDLINE
DN 92276334 PubMed ID: 1592810
TI High information conservation implies that at least three **proteins** bind independently to F plasmid incD repeats.
AU Herman N D; Schneider T D
CS Laboratory of Mathematical Biology, National Cancer Institute, Frederick Cancer Research and Development Center, Maryland 21702-1201.
SO JOURNAL OF BACTERIOLOGY, (1992 Jun) 174 (11) 3558-60.

Journal code: 2985120R. ISSN: 0021-9193.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-M79367; GENBANK-M79368; GENBANK-M79369; GENBANK-M79370; GENBANK-M79371; GENBANK-M79372; GENBANK-M81963; GENBANK-M87491; GENBANK-M87492; GENBANK-X04619; GENBANK-X13141
EM 199207
ED Entered STN: 19920710
Last Updated on STN: 19920710
Entered Medline: 19920701
AB The 12 incD repeats in the F plasmid each contain about 60 **bits** of information, which is three times the amount of conservation that a single **protein** would need to distinguish the repeats from the rest of the Escherichia coli genome. This is the first reported discovery of a case of threefold excess information, and it implies that at least three **proteins** bind independently to the repeats. In support of this observation, other workers have shown that three **polypeptides** bind to this region, but only one, SopB, is known to bind independently of other factors. Identification of the other two **proteins** should help us to understand the mechanism of plasmid partitioning during cell division.

=> d his

(FILE 'HOME' ENTERED AT 17:01:59 ON 04 SEP 2002)

FILE 'MEDLINE, BIOSIS' ENTERED AT 17:02:09 ON 04 SEP 2002
L1 128 S SEQUENCE AND BITS
L2 67 S L1 AND (DNA OR PROTEIN OR POLYPEPTIDE OR POLYNUCLEOTIDE OR AM
L3 42 DUPLICATE REMOVE L2 (25 DUPLICATES REMOVED)

FILE 'STNGUIDE' ENTERED AT 17:13:38 ON 04 SEP 2002

FILE 'MEDLINE, BIOSIS' ENTERED AT 17:15:57 ON 04 SEP 2002

=> s sequence and compression
L4 1036 SEQUENCE AND COMPRESSION

=> s l4 and (dna or protein or polypeptide or nucleotide or amino acid)
L5 298 L4 AND (DNA OR PROTEIN OR POLYPEPTIDE OR NUCLEOTIDE OR AMINO ACID)

=> s l5 and (bit or bits)
L6 10 L5 AND (BIT OR BITS)

=> duplicate remove l6
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE? Y/(N):n
PROCESSING COMPLETED FOR L6
L7 6 DUPLICATE REMOVE L6 (4 DUPLICATES REMOVED)

=> d 1-6 bib ab

L7 ANSWER 1 OF 6 MEDLINE
AN 2001646608 MEDLINE
DN 21557032 PubMed ID: 11700586
TI Biological **sequence compression** algorithms.
AU Matsumoto T; Sadakane K; Imai H
CS Department of Information Science, University of Tokyo,7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.. toshikom@is.s.u-tokyo.ac.jp
SO GENOME INFORMATICS SERIES, (2000) 11 43-52.
Journal code: 9717234. ISSN: 0919-9454.

CY Japan
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200112
ED Entered STN: 20011112
Last Updated on STN: 20020124
Entered Medline: 20011231
AB Today, more and more **DNA sequences** are becoming available. The information about **DNA sequences** are stored in molecular biology databases. The size and importance of these databases will be bigger and bigger in the future, therefore this information must be stored or communicated efficiently. Furthermore, **sequence compression** can be used to define similarities between biological **sequences**. The standard **compression** algorithms such as gzip or compress cannot compress **DNA sequences**, but only expand them in size. On the other hand, CTW (Context Tree Weighting Method) can compress **DNA sequences** less than two **bits** per symbol. These algorithms do not use special structures of biological **sequences** . Two characteristic structures of **DNA sequences** are known. One is called palindromes or reverse complements and the other structure is approximate repeats. Several specific algorithms for **DNA sequences** that use these structures can compress them less than two **bits** per symbol. In this paper, we improve the CTW so that characteristic structures of **DNA sequences** are available. Before encoding the next symbol, the algorithm searches an approximate repeat and palindrome using hash and dynamic programming. If there is a palindrome or an approximate repeat with enough length then our algorithm represents it with length and distance. By using this preprocessing, a new program achieves a little higher **compression** ratio than that of existing **DNA** -oriented **compression** algorithms. We also describe new **compression** algorithm for **protein sequences**.

L7 ANSWER 2 OF 6 MEDLINE DUPLICATE 1
AN 1999238185 MEDLINE
DN 99238185 PubMed ID: 10223669
TI Significantly lower entropy estimates for natural **DNA sequences**.
AU Loewenstern D; Yianilos P N
CS NEC Research Institute, Princeton, New Jersey 08540, USA.. davel@research.nj.nec.com
SO JOURNAL OF COMPUTATIONAL BIOLOGY, (1999 Spring) 6 (1) 125-42. Journal code: 9433358. ISSN: 1066-5277.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199906
ED Entered STN: 19990618
Last Updated on STN: 19990618
Entered Medline: 19990607
AB If **DNA** were a random string over its alphabet A, C, G, T , an optimal code would assign two **bits** to each **nucleotide**. **DNA** may be imagined to be a highly ordered, purposeful molecule, and one might therefore reasonably expect statistical models of its string representation to produce much lower entropy estimates. Surprisingly, this has not been the case for many natural **DNA sequences**, including portions of the human genome. We introduce a new statistical model (**compression** algorithm), the strongest reported to date, for naturally occurring **DNA sequences**. Conventional techniques code a **nucleotide** using only slightly fewer **bits** (1.90) than one obtains by relying only on the frequency statistics of individual **nucleotides** (1.95). Our method in some

cases increases this gap by more than fivefold (1.66) and may lead to better performance in microbiological pattern recognition applications. One of our main contributions, and the principle source of these improvements, is the formal inclusion of inexact match information in the model. The existence of matches at various distances forms a panel of experts which are then combined into a single prediction. The structure of this combination is novel and its parameters are learned using Expectation Maximization (EM). Experiments are reported using a wide variety of **DNA sequences** and compared whenever possible with earlier work. Four reasonable notions for the string distance function used to identify near matches, are implemented and experimentally compared. We also report lower entropy estimates for coding regions extracted from a large collection of nonredundant human genes. The conventional estimate is 1.92 **bits**. Our model produces only slightly better results (1.91 **bits**) when considering **nucleotides**, but achieves 1.84-1.87 **bits** when the prediction problem is divided into two stages: (i) predict the next **amino acid**-based on inexact **polypeptide** matches, and (ii) predict the particular codon. Our results suggest that matches at the **amino acid** level play some role, but a small one, in determining the statistical structure of nonredundant coding **sequences**.

L7 ANSWER 3 OF 6 MEDLINE DUPLICATE 2
AN 93189555 MEDLINE
DN 93189555 PubMed ID: 8446578
TI Sequencing two **DNA** templates in five channels by digital **compression**.
AU Nelson M; Zhang Y; Steffens D L; Grabherr R; Van Etten J L
CS Department of Biochemistry and Molecular Biology, University of Chicago, IL 60637.
NC GM-32441 (NIGMS)
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1993 Mar 1) 90 (5) 1647-51.
Journal code: 7505876. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199304
ED Entered STN: 19930416
Last Updated on STN: 19980206
Entered Medline: 19930408
AB By applying algebraic coding methods to the Sanger dideoxynucleotide procedure, **DNA sequences** of two templates can be determined simultaneously in only five reactions and data channels. A 5:2 data **compression** is accomplished by instantaneous source coding of **nucleotide sequence** pairs into one set of 5-**bit** block codes. A general algebraic expression, 2n-1 > or = 4f, describes conditions under which f **DNA** templates can be sequenced using n channels. Such **compression** sequencing is accurate and efficient, as demonstrated by manual 35S autoradiographic detection and automated on-line analysis using fluorescent-labeled primers. Symmetric 5:2 **compression** is especially useful when comparing two closely related **sequences**.

L7 ANSWER 4 OF 6 MEDLINE DUPLICATE 3
AN 94005759 MEDLINE
DN 94005759 PubMed ID: 8402207
TI Discovering simple **DNA sequences** by the algorithmic significance method.
AU Milosavljevic A; Jurka J
CS Linus Pauling Institute of Science and Medicine, Palo Alto, CA 94306.
SO COMPUTER APPLICATIONS IN THE BIOSCIENCES, (1993 Aug) 9 (4) 407-11.
Journal code: 8511758. ISSN: 0266-7061.

CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199311
ED Entered STN: 19940117
Last Updated on STN: 19980206
Entered Medline: 19931119
AB A new method, 'algorithmic significance', is proposed as a tool for discovery of patterns in **DNA sequences**. The main idea is that patterns can be discovered by finding ways to encode the observed data concisely. In this sense, the method can be viewed as a formal version of the Occam's Razor principle. In this paper the method is applied to discover significantly simple **DNA sequences** . We define **DNA sequences** to be simple if they contain repeated occurrences of certain 'words' and thus can be encoded in a small number of **bits**. Such definition includes minisatellites and microsatellites. A standard dynamic programming algorithm for data **compression** is applied to compute the minimal encoding lengths of **sequences** in linear time. An electronic mail server for identification of simple **sequences** based on the proposed method has been installed at the Internet address pythia/anl.gov.

L7 ANSWER 5 OF 6 MEDLINE
AN 96038977 MEDLINE
DN 96038977 PubMed ID: 7584347
TI Discovering **sequence** similarity by the algorithmic significance method.
AU Milosavljevic A
CS Biological and Medical Research Division, Argonne National Laboratory, Illinois 60439-4833, USA.
SO ISMB, (1993) 1 284-91.
Journal code: 9509125.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199512
ED Entered STN: 19960124
Last Updated on STN: 19980206
Entered Medline: 19951205
AB The minimal-length encoding approach is applied to define concept of **sequence** similarity. A **sequence** is defined to be similar to another **sequence** or to a set of keywords if it can be encoded in a small number of **bits** by taking advantage of common subwords. Minimal-length encoding of a **sequence** is computed in linear time, using a data **compression** algorithm that is based on a dynamic programming strategy and the directed acyclic word graph data structure. No assumptions about common word ("k-tuple") length are made in advance, and common words of any length are considered. The newly proposed algorithmic significance method provides an exact upper bound on the probability that **sequence** similarity has occurred by chance, thus eliminating the need for any arbitrary choice of similarity thresholds. Preliminary experiments indicate that a small number of keywords can positively identify a **DNA sequence**, which is extremely relevant in the context of partial sequencing by hybridization.

L7 ANSWER 6 OF 6 MEDLINE DUPLICATE 4
AN 92339043 MEDLINE
DN 92339043 PubMed ID: 1378774
TI Natural **sequence** code representations for **compression** and rapid searching of human-genome style databases.
AU Robson B; Greaney P J
CS Department of Molecular Biology, University of Manchester, UK.

SO COMPUTER APPLICATIONS IN THE BIOSCIENCES, (1992 Jun) 8 (3) 283-9.
Journal code: 8511758. ISSN: 0266-7061.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199208
ED Entered STN: 19920911
Last Updated on STN: 19960129
Entered Medline: 19920825
AB Numeric descriptions ('bio-informatic descriptions') of **amino acid** residues have been developed which will be of value whenever the quality and quantity of information in very large (i.e. 'human genome style') gene and **protein sequences** is to be compared or manipulated. These codes are as natural as possible by our criteria (the same principles could be used in revision of the criteria). In particular, in storing and searching large amounts of **sequence** data, natural codes--which relate to the properties of **amino acids**--can be combined with existing fast-search algorithms but introduce several advantages. The code can be assigned such that sub-selection of **bits** leads to compressed databases with residues defined less specifically, by classes of properties. The most compressed representation leads to the specification of a residue as polar or non-polar, while the most extended representation used at present also allows specification of, for example, glyco-asparagine and phosphoserine. Preliminary studies on both a supercomputer and smaller machines suggest a 'worst-case' speeding of approximately 4.5-fold. For more intelligent searching, coding extensions mixed with the basic **sequence** data give the **sequence** data some of the character of a computer program.

=> s (context tree weighting method)
L8 1 (CONTEXT TREE WEIGHTING METHOD)

=>

---Logging off of STN---

=>
Executing the logoff script...

=> LOG Y

COST IN U.S. DOLLARS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	10.41	31.93

STN INTERNATIONAL LOGOFF AT 17:22:56 ON 04 SEP 2002